2011

METHYLATION AND PATTERNS OF EVOLUTION IN IMPRINTED GENES IN THE MOUSE

Tori A. LaFleur
Northern Michigan University

Follow this and additional works at: https://commons.nmu.edu/theses

Recommended Citation
https://commons.nmu.edu/theses/431

This Open Access is brought to you for free and open access by the Student Works at NMU Commons. It has been accepted for inclusion in All NMU Master's Theses by an authorized administrator of NMU Commons. For more information, please contact kmcdonou@nmu.edu, bsarjean@nmu.edu.
METHYLATION AND PATTERNS OF EVOLUTION IN IMPRINTED GENES IN THE MOUSE

By

Tori A. LaFleur

THESIS

Submitted to
Northern Michigan University
In partial fulfillment of the requirements
For the degree of

MASTER OF SCIENCE

Graduate Studies Office

2011
This thesis by Tori A. LaFleur is recommended for approval by the student’s thesis committee in the Department of Biology and by the Dean of Graduate Studies.

____________________________________________________
Committee Chair: Dr. Katherine Teeter
Date

____________________________________________________
First Reader: Dr. John Rebers
Date

____________________________________________________
Second Reader: Dr. Alec Lindsay
Date

____________________________________________________
Department Head: Dr. Patrick Brown
Date

____________________________________________________
Dean of Graduate Studies: Dr. Terry Seethoff
Date
In order to catalog your thesis properly and enter a record in the OCLC international bibliographic data base, Olson Library must have the following requested information to distinguish you from others with the same or similar names and to provide appropriate subject access for other researchers.

NAME: LaFleur, Tori Anne

DATE OF BIRTH: November 6, 1986
ABSTRACT

METHYLATION AND PATTERNS OF EVOLUTION IN IMPRINTED GENES IN THE MOUSE

By

Tori A. LaFleur

Genomic imprinting presents a fascinating challenge for geneticists. In addition to the rarity of imprinted loci in the genome, the mechanism by which imprinted genes are functionally silenced in offspring was identified as methylation at the fifth carbon of cytosine in cytosine-guanine dinucleotides only recently. A more comprehensive understanding of how this mechanism operates in humans is critical to ascertaining the effects of its deregulation on human hereditary syndromes and in cancer pathways. More information is required about the deregulation of imprinted loci, also termed loss of imprinting (LOI), in mouse species of the genus Mus, the most widely used organisms in biomedical research. This study investigated LOI resulting from hybridization between Mus musculus musculus and M. m. domesticus by examining loss of 5-methylcytosine nucleotides in the control regions of four imprinted genes. Statistical analysis was used to determine whether natural selection imposes greater pressure on imprinted loci as compared to neutral genes.
Copyright by
Tori A. LaFleur
2011
DEDICATION

I would like to thank my family, Ellwood, Lorraine, and James; my best friend, Jacqueline; and my partner, Adam, for their unwavering confidence in me.
ACKNOWLEDGMENTS

I am grateful to my adviser, Dr. Katherine Teeter, for providing all of the time, knowledge, and encouragement I required; my committee members, Dr. Alec Lindsay and Dr. John Rebers, for their enthusiasm and expertise; Dr. Bret Payseur for providing required tissue samples; and the Northern Michigan University Department of Biology for their support.
PREFACE

The cost of the research covered in this project has been underwritten by grants from the Northern Michigan University College of Graduate Studies and Office of Continuing Education and from the laboratory of Dr. Katherine Teeter.

This thesis follows the format prescribed by the journal Cell and by Northern Michigan University’s Department of Biology. The usual convention of naming human genes in uppercase italics and non-human genes in lowercase italics was employed. When describing genetic crosses, the species or subspecies of the female parent was listed first.
TABLE OF CONTENTS

List of Tables...........................................................................................................(viii)

List of Figures...........................................................................................................(ix)

List of Symbols and Abbreviations...........................................................................(x)

Introduction.............................................................................................................1

Chapter One: Literature Review
  1.1 The Origin of Genomic Imprinting.................................................................2
  1.2 Methylation of CpG Dinucleotides.................................................................6
  1.3 Mus Species in Biomedical Research .............................................................9
  1.4 Parent-of-Origin Effects and Hybrid Dysgenesis ...........................................11
  1.5 Genomic Imprinting and Mammalian Speciation ..........................................12
  1.6 Molecular Evolution of Imprinted Loci ........................................................16
  1.7 Medical and Environmental Considerations ...............................................18

Chapter Two: Methylation in Mouse Hybrids
  2.1 Bisulfite Sequencing of Imprinted Loci........................................................21
    2.1.1 H19....................................................................................................23
    2.1.2 Igf2r.................................................................................................23
    2.1.3 Rasgrf1............................................................................................25
    2.1.4 Snrpn...............................................................................................25
  2.2 Methods and Materials..................................................................................26
  2.3 Results..........................................................................................................28
    2.3.1 H19....................................................................................................28
    2.3.2 Igf2r.................................................................................................32
    2.3.3 Rasgrf1............................................................................................35
    2.3.4 Snrpn...............................................................................................38
  2.4 Discussion......................................................................................................42
    2.4.1 Conservation of CpG Methylation at Igf2r............................................42
    2.4.2 Loss of CpG Methylation at H19, Rasgrf1, and Snrpn..........................43
    2.4.3 Implications for Mammalian Speciation..............................................44
    2.4.4 Genomic Imprinting and LOI in Mus Intersubspecific Hybrids..............47
    2.4.5 Hybrid Sterility and Haldane’s Rule....................................................48

Chapter Three: Selective Pressures at Imprinted Loci
  3.1 Natural Selection and Genomic Imprinting..................................................50
  3.2 The $d_s/d_0$ ratio in Molecular Evolution......................................................51
  3.3 Bioinformatic Analysis of Molecular Evolution.........................................54
  3.4 Methods and Materials.................................................................................56
  3.5 Results..........................................................................................................57
TABLE OF CONTENTS

3.6 Discussion.........................................................................................................................62
  3.6.1 Site-Specific Positive Selection.................................................................63
  3.6.2 Functions of *Rasgrf1* in Mammalian Physiology.............................65
  3.6.3 Overall Absence of Positive Selection at Imprinted Loci.............68

Chapter Four: Summary and Conclusions.................................................................72

References...............................................................................................................................76

Appendix A..............................................................................................................................96

Appendix B..............................................................................................................................99
LIST OF TABLES

Table 1: Conserved clusters between the human and mouse genomes……………………4
Table 2: Summary of CpG methylation at four imprinted loci…………………………41
Table 3: Mann-Whitney U-test statistics for comparison of $d_S/d_S$ ratios……………60
Table 4: Primer sequences for amplifying negative controls…………………………96
Table 5: Mouse and rat orthologs of imprinted loci…………………………………….99
Table 6: Mouse and rat orthologs of non-essential loci……………………………………99
Table 7: Mouse and rat orthologs of immune loci……………………………………100
Table 8: Mouse and rat orthologs of essential loci………………………………………100
Table 9: Rates of molecular evolution at mouse and rat imprinted loci…………………101
Table 10: Rates of molecular evolution at mouse and rat non-essential loci………………101
Table 11: Rates of molecular evolution at mouse and rat essential loci…………………102
Table 12: Rates of molecular evolution at mouse and rat immune loci…………………102
LIST OF FIGURES

Figure 1: Igf2 imprinting in mice ................................................................. 2
Figure 2: Establishment of DNA methylation in the germ line ................. 8
Figure 3: Bisulfite-induced deamination of cytosine to uracil .................. 22
Figure 4: Comparison of CpG methylation of the H19 DMR .................. 29
Figure 5: Detection of CpG methylation at one allele .......................... 31
Figure 6: Comparison of CpG methylation of the Igf2r DMR ............... 33
Figure 7: Comparison of CpG methylation of the Rasgrf1 DMR ............ 36
Figure 8: Comparison of CpG methylation of the Snrpn DMR ............... 39
Figure 9: Number of synonymous and nonsynonymous substitutions ....... 58
Figure 10: Mean ratios of nonsynonymous and synonymous substitution rates ........ 59
Figure 11: Sliding window analysis of Rasgrf1 loci in four species ............ 61
Figure 12: Zymo Research methylated mouse gDNA control ............... 98
LIST OF SYMBOLS AND ABBREVIATIONS

ART: assisted reproductive technology
AS: Angelman syndrome
BLAST: Basic Local Alignment Search Tool
BWS: Beckwith-Wiedemann syndrome
CpG: cytosine-guanine dinucleotide
CGI: CpG island
CTCF: CCTC binding factor (transcription factor)
$d_N$: rate of nonsynonymous amino acid substitutions
$d_S$: rate of synonymous amino acid substitutions
$d_N/d_S$: ratio of nonsynonymous to synonymous substitution ($\omega$)
DMD: differentially methylated domain
DMR: differentially methylated region
DNMT: DNA methyltransferase
d.p.c.: days postcoitum
ERK: extracellular signal-regulated kinase
gDNA: genomic DNA
$H_0$: null hypothesis
$H_A$: alternative hypothesis
ICR: imprinting control region
IVF: in vitro fertilization
LOI: loss of imprinting
NMDAR: \textit{N}-methyl-D-aspartate glutamate receptor

\textit{q}: region located on long arm of chromosome

\textit{p}: region located on short arm of chromosome

PCR: polymerase chain reaction

PGC: primordial germ cell

PWS: Prader-Willi syndrome

RT-PCR: reverse transcriptase PCR

SNP: single nucleotide polymorphism

SRS: Silver-Russell syndrome
INTRODUCTION

Genomic imprinting, also referred to as parental imprinting, is a reversible form of epigenetic control. Epigenetic modifications are mechanisms that allow for heritable changes of gene expression without alteration of the actual genetic sequence (Haig, 2002). Genomic imprinting marks the copy of a gene inherited from either the mother or father with a chemical group, which leads to silencing of transcription from that copy in the offspring. Due to the influence of epigenetics, it is possible to alter—sometimes drastically—the functions of the genome while the sequence remains unchanged (Jablonka and Lamb, 2002). Investigations of imprinting are frequently conducted using laboratory mice of several subspecies of M. musculus. As premier mammalian models for biomedical research, Mus species offer tremendous insight into the human genome and the molecular mechanisms underlying its regulation.

As a consequence of genomic imprinting, the form of a gene, or allele, expressed in the offspring depends on the parent from whom it was inherited. Though each parent does contribute an allele of the same gene, which is found at the same locus on the chromosome, only one of the two alleles is expressed in the offspring. The allele which is functionally silenced in the offspring is termed imprinted (Haig, 2000; Pulford et al., 1999). Like mice, humans experience deleterious effects from disruption of imprinting, including the development of some cancers (Paulsen and Ferguson-Smith, 2001; Pulford et al., 1999). Research in genomic imprinting is therefore an important step in improving our identification and understanding of a number of human pathologies.
1.1 The Origin of Genomic Imprinting

Classical genetics assumes that each parent contributes equally to the traits expressed in their progeny. As a consequence of imprinting, this basic principle is violated: the parental complements become unequal, with the expression of maternally- and paternally-derived genes in offspring differing due to the presence of epigenetic marks (represented in Figure 1, below). The resulting functional haploidy is often associated with disease, as there is no protection against harmful recessive mutations (Dolinoy et al., 2007; Hore et al., 2007).

**Figure 1: Igf2 imprinting in mice.** Following the inheritance of both maternal and paternal alleles by the offspring, silencing results in monoallelic expression from the Igf2 locus.
A number of imprinted genes demonstrate tissue-specific patterns. In more complex cases there may be only a quantitative difference in expression (Burt and Trivers, 1998). Imprinted genes play a role in a variety of cellular and physiological processes, including cell proliferation, fetal and placental growth, and establishment of adult behavioral patterns (Falls et al., 2002).

Recent estimates place the number of imprinted genes in the mouse genome at fewer than 100 (Williamson et al., 2011). Imprinting is controlled by regions generally ranging from 1-5 kb in size, which are referred to as differentially methylated regions (DMR) or differentially methylated domains (DMD). Within each DMR, one parental allele will be imprinted. DMRs vary widely in location: while some are found within promoter sequences or introns, others may be located upstream of the promoter region (Reinhart et al., 2002). For over a decade, DMRs have been known to be crucial to the maintenance of imprinted expression (Bielinska et al., 2000; Thorvaldsen et al., 1998).

Imprinting is the most common cause of parent-of-origin effects in mammalian genomes (Davies et al., 2005; Vrana et al., 2007). Primary parental imprints are believed to be established during gametogenesis, when both genomes are separate. Imprints must then survive the reprogramming events that occur in the preimplantation embryo following fertilization (Ideraabdullah et al., 2008). Inheritance through the mitotic divisions of fetal development requires the response of trans-acting genes in the offspring, which are responsible for reading and maintenance of the parental imprint (Burt and Trivers, 1998).
Like that of the mouse, the human genome demonstrates clustering of imprinted loci, which are also referred to as *imprinted domains* (Beaudet and Jiang, 2002). Most clusters between the mouse and human genomes are highly conserved (see Table 1), making the mouse an especially useful and relevant species in the study of imprinting disorders (Hall, 1990). Given that the majority of imprinted genes identified are located within clusters, it is possible that one imprinting control region (ICR) is able to affect methylation and gene expression over multiple loci, making it the primary *cis*-acting element controlling function (Pearsall *et al*., 1999). Imprinted regions are also associated with asynchronous DNA replication, suggesting that the timing of replication is differentially regulated on the two parental chromosomes (Kitsberg *et al*., 1993).

**Table 1: Conserved clusters of imprinted genes between the human and mouse genomes.** Approximately half of imprinted loci are located within these five clusters (Glaser and Morison, 2009).

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Chr. (Mouse)</th>
<th>Chr. (Human)</th>
<th>Representative Genes</th>
<th>Investigators</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kcnq10t1</td>
<td>7</td>
<td>11p15</td>
<td><em>Kcnq1, Cdkn1c</em></td>
<td>Lin <em>et al.</em> (2003)</td>
</tr>
<tr>
<td>Dlk1</td>
<td>12</td>
<td>14q32</td>
<td><em>Dlk1, Gtl2</em></td>
<td>Takada <em>et al.</em> (2002)</td>
</tr>
<tr>
<td>Peg3</td>
<td>7</td>
<td>19q13</td>
<td><em>Prkcg, Peg3, Znf134, Znf132, Znf42</em></td>
<td>Kim <em>et al.</em> (1997)</td>
</tr>
<tr>
<td>Peg10</td>
<td>6</td>
<td>7q21</td>
<td><em>Peg10, Sgce, Col1A2</em></td>
<td>Ono <em>et al.</em> (2001)</td>
</tr>
</tbody>
</table>
The hypothesis that imprinting evolved as a means to inactivate foreign or parasitic DNA sequences has led to a number of fascinating studies (Pulford et al., 1999). Defense against parasitic sequences such as retrotransposons and retroviruses, which are estimated to comprise at least 35% of the human genome, is believed to depend on methylation of the CpG dinucleotides abundant in these regions (Bestor, 1998). Differential methylation of CpG dinucleotides within maternally- and paternally-inherited alleles remains the leading candidate in the search for the mechanism underlying imprinted expression.

Other researchers describe the evolution of genomic imprinting as a matter of supply and demand in which most imprinted genes interact between the placenta and fetus in order to regulate placental transfer capacity (Angiolini et al., 2006; Reik et al., 2003). Past studies have also suggested that imprinting in female mice may be responsible for the preferential inactivation of the paternally-derived X chromosome, deregulation of which is implicated in abnormal phenotypes of complex behavioral characteristics such as maternal nurturing (Davies et al., 2005; Li et al., 1999).

Whatever its origin, genomic imprinting continues to influence the development and inheritance of a diverse array of mammalian pathologies (see Medical Considerations, Section 1.7). These include male infertility, pre-eclampsia, behavioral disorders, and tumorigenesis in multiple organs (Falls et al., 1999; Filipponi and Feil, 2009; Nakao and Sasaki, 1996; Rodenhiser and Mann, 2006). Investigating the molecular mechanisms employed to impose epigenetic control has become an increasingly important field of research (Wolffe and Matzke, 1999).
1.2 Methylation of CpG Dinucleotides

Methylation of cytosine is observed at the fifth carbon atom of cytosine in cytosine-guanine dinucleotides within imprinted sequences (Bird, 2002). Modification of the DNA sequence with methyl groups satisfies the criteria required for an imprinting mechanism, including the ability to silence transcription of genes, to be propagated stably to daughter cells, and to be removed in the presence of active demethylases in order to establish or renew transcription (Tycko and Morison, 2002). The fact that imprinting effects have not been observed in species whose DNA cannot be modified by methylation offers further evidence supporting cytosine methylation as the regulatory mechanism of imprinting (Bestor, 1998). Different degrees of methylation may therefore be responsible for the silencing of one of two parental alleles in the offspring, resulting in imprinted expression patterns.

The addition of methyl groups to the carbon-5 position of cytosine nucleotides following DNA synthesis modifies the appearance of the major groove of the DNA helix. For imprinting to be inherited through mitotic divisions in somatic cell lineages, CpG dinucleotides must have one methylcytosine segregated to each daughter DNA duplex following semiconservative replication. Each hemimethylated dinucleotide is then remethylated by DNA methyltransferase (DNMT) 1 (Bestor, 1998; Wolffe and Matzke, 1999). Methylation is believed to alter chromatin structure at CpG-rich promoters such that DNA binding proteins are unable to initiate transcription or can do so only at a significantly lower rate (Jones and Takai, 2001). The formation of transcriptionally protected chromatin is termed an insulating mechanism and is thought to depend on the
activity of the DNA-binding protein CCCTC binding factor (CTCF), a potent transcriptional repressor that acts preferentially on CG-rich DNA (Wolffe, 2000).

It has been known for some time that mouse embryos lacking the maintenance enzyme Dnmt1 die at approximately day E9.5 of development (Schmidt et al., 2000). Further, evidence suggests that Dnmt1 acts preferentially on a hemimethylated substrate (Bestor, 2000). More recent studies have indicated that the de novo methyltransferases Dnmt3a and Dnmt3b may be responsible for initiating imprinting and altering methylation status during embryonic development (Kaneda et al., 2004; Okano et al., 1999).

While imprinting is stably maintained in somatic cells, it has been shown that erasure and reestablishment occur in germ cell lines to reflect the sex of the individual (Reik and Walter, 2001). Primordial germ cells (PGCs) demonstrate removal of methyl groups at CpG dinucleotides in DMRs and biallelic states of gene expression from days E11.5 - E13.5 of development, indicating loss of imprinting (Lee et al., 2002). See Figure 2 for an overview of this process. Remethylation is believed to occur at the prospermatogonia stage in the male germ line (E15) and later in the female germ line, following birth and during growth of the oocytes (Reik et al., 2001).
Figure 2: Establishment of DNA methylation in the germ line. Following fertilization, imprinting is stably maintained in the somatic cell lineage. In contrast, germ cells undergo demethylation of the full genome followed by later remethylation. The figure is adapted from Rakyan et al. (2001) and depicts a hypothetical paternally imprinted gene.

Clusters of CpG dinucleotides, termed CpG islands (CGI), generally contain the highest density of CpGs in normal tissues and are frequently found in the promoter or regulatory regions required for gene expression (Bird and Wolffe, 1999; Esteller, 2006). CpG dinucleotides are underrepresented in the mammalian genome, likely due to the risk of spontaneous deamination (see Molecular Evolution of Imprinted Loci, Chapter 1.6) (Dolinoy et al., 2007). Control regions for these clusters have been tentatively identified as the DMRs and ICRs (Yotova et al., 2008). ICRs have been associated with antisense transcripts capable of silencing upstream genes through RNA interference; occurring primarily in the female germ line, this process ensures consistent monoallelic expression
of imprinted genes (Surani, 2001). Following methylation of cytosines, promoters of imprinted genes are observed to be stably silent (Bird, 2002).

1.3  *Mus* Species in Biomedical Research

McGrath and Solter (1984) used nuclear transplantation to create mouse embryos containing only one of the two sets of parental chromosomes. Following the removal of a single pronucleus, they introduced a second from another embryo, constructing diploid gynogenetic and androgenetic embryos. They found that embryogenesis did not proceed normally. Despite the fact that the embryos contained the normal amount of genomic DNA, gynogenetic embryos did not proceed past the early somite stage before dying due to growth retardation and undeveloped extraembryonic tissues; androgenetic embryos arrested at an even earlier stage of development while their extraembryonic tissues became grossly overgrown. Only a diploid genome derived from both maternal and paternal complements was found to capable of directing the growth of a viable embryo.

Surani *et al.* (1984) came to similar conclusions when they introduced a second female pronucleus to mouse oocytes, transferred them to the oviducts, and allowed the embryos to implant. Only control embryos which contained male pronuclei developed to term; those with two female pronuclei were unable to develop normal trophoblasts and extraembryonic membranes. Essentially, the two parental genomes were found to be nonequivalent, a conclusion McGrath and Solter found surprising given the number of species in which parthenogenotes are viable. Both groups were startled to find that genes appeared to function differently in the embryos depending on whether they originated from the mother or father, results that suggested the presence of a heritable imprint in the
germ line which was then manifested in the somatic cells of the offspring (Varmuza, 1993).

*Insulin-like growth factor 2 (Igf2)* was the first imprinted locus identified in *Mus* and is exclusively expressed from the paternal allele in most tissues (Vrana et al., 2007). *Insulin-like growth factor 2 receptor (Igf2r)* demonstrates the opposite pattern of expression: it is exclusively expressed from the maternal allele. *Igf2* expression promotes growth by binding to the insulin-like growth factor 1 receptor or the insulin receptor (Hawkes and Kar, 2004), while *Igf2r* expression leads to degradation of Igf2 protein, inhibiting growth (Haig, 1997). The receptor plays a role in clearing Igf2 from the circulation by sequestering the protein once it is internalized and exposing it to lysosomal enzymes (Pavelić et al., 2007; Rezgui et al., 2009). Imprinted expression may also demonstrate tissue specificity. The normal paternal expression of *Igf2* is altered in the choroid plexus and leptomeninges, where biallelic expression is seen (Mochizuki et al., 1996).

Clustering of human imprinted genes is observed, with many clusters appearing to correlate with those found in the mouse genome (Table 1). Murine clusters often combine several protein-coding genes with at least one non-coding RNA; the possibility of similar structures is currently being explored in other mammals (Edwards and Ferguson-Smith, 2007; Yotova et al., 2008). The cluster associated with Prader-Willi and Angelman syndromes, located on human chromosome 15, is syntenic with the central region of mouse chromosome 7, containing a number of imprinted loci (Nicholls et al., 1998).
1.4 Parent-of-Origin Effects and Hybrid Dysgenesis

In interspecies hybrids, expression of imprinted genes has been shown to differ between reciprocal crosses. The genetic phenomenon in which a phenotype varies depending on whether the father or mother contributed the allele is a parent-of-origin effect, resulting in offspring who are not identical as would be predicted by a Mendelian model. Though mice are frequently used for mammalian hybridization studies, parent-of-origin effects have been investigated in a number of other species. For example, crosses between horses and donkeys have been intensively studied in an attempt to understand why a female horse mated with a male donkey produces a mule, while a male horse mated with a female donkey produces a hinny (Allen et al., 1993). Not all imprinted genes in the hybrid offspring of two species demonstrate the same pattern, suggesting that different loci are under separate control to some extent (Burt and Trivers, 1998). The abnormalities occurring from interspecies hybridization are collectively referred to as hybrid dysgenesis effects and are believed to be due in part to deregulation of imprinting (Duselis and Vrana, 2007).

Mouse hybrid models have shown a number of fascinating hybrid dysgenesis phenotypes. Those performed under laboratory conditions have generally been between *M. m. musculus* and *M. spretus* and resulted in placental dysplasia in female hybrids (Zechner et al., 1996). Occurrences of LOI for *Igf2, Igf2r, H19, small nuclear ribonucleoprotein polypeptide N (Snrpn), growth factor receptor-bound protein 10 (Grb10), Mash2, and mesoderm-specific transcript protein (Mest)*, also referred to as paternally-expressed gene 1 (Peg1), have been documented in crosses between the mouse species *Peromyscus maniculatus* and *P. polionotus*, indicating that imprinting signals
from one species may not be recognized in the other (Jirtle et al., 2000; Vrana et al., 1998). Differential expression due to imprinting effects is thought to be the cause of hybrid dysgenesis in many mouse species. Because small hybrid animals would fare so poorly in the wild and overgrowth in *Peromyscus* hybrids has been documented as lethal, LOI is a potential source of postzygotic incompatibilities and is thus heavily implicated in speciation (see Genomic Imprinting and Mammalian Speciation, Chapter 1.5) (Vrana et al., 1998, 2000; Zeh and Zeh, 2000).

In *Peromyscus* interspecific hybrids, abnormal methylation patterns appear to correlate with deregulation of imprinting in multiple genes. Recent studies suggest that many genes demonstrating altered methylation state and LOI in *Peromyscus* hybrids are also deregulated in *Mus* hybrids (Shi et al., 2004). Changes in fetal growth and placental size are among the most widely studied abnormalities in *Peromyscus*, which have been attributed to LOI of the genes *Igf2* and *Igf2r* (Reik and Walter, 2001). Although it was once argued that widespread LOI does not occur in *Mus* hybrids as it does in *Peromyscus* hybrids (Vrana et al., 1998), more recent studies have clearly refuted this (Shi et al., 2004). Imprinted genes that are suspected to suffer LOI in both *Peromyscus* and *Mus* hybrids include *Igf2r*, *H19*, *Ras protein-specific guanine nucleotide-releasing factor 1* (*Rasgrf1*), and *Snrpn.

1.5 Genomic Imprinting and Mammalian Speciation

Genomic imprinting is critical in embryogenesis, a consequence of the role of imprinted genes in placental physiology and fetal development. *In vitro* manipulations performed to obtain strictly androgenetic or gynogenetic embryos have confirmed the
need for both maternal and paternal chromosomal complements in order to produce viable embryos (e.g., Surani et al., 1984). Functional alleles must be present from each parent for the establishment of correct expression patterns. Offspring with two maternal copies often display growth retardation and placental abnormalities, whereas those with two paternal copies show enhanced fetal and placental growth and remain abnormally large into adulthood (Cattanach et al., 2004).

Interestingly, many imprinted genes involved in fetal growth serve opposing functions: paternally-expressed Igf2 and Mest/Peg1 promote growth, while maternally-expressed Igf2r and Cdkn1c inhibit it (Eggerman et al., 2008). In addition to Igf2 and Igf2r, the genes Peg3, Zim1, and Usp29 have been implicated in increased fetal growth and placental size as a result of paternal control and maternal suppression during gestation. These data are consistent with the parental conflict hypothesis of Moore and Haig, which posits that expression of paternal alleles will increase the cost of bearing offspring to the mother while expression of maternal alleles counteracts this demand (Cattanach et al., 2004). This conflict is thought to be the result of polyandrous mating in mammals and the unequal investments of the mother and father in their offspring (Jirtle and Skinner, 2007). The result of this model is that the offspring of one male are in competition for nutrients with the offspring of another male birthed by the same female, their half-siblings (Haig, 2002). Maternal and paternal alleles would thus favor different levels of gene product, an interplay which Haig (1997) termed parental antagonism.

Some researchers have suggested that this intralocus sexual conflict between paternal and maternal germ lines has selected for genomic imprinting during mammalian
evolution. Killian et al. (2001) reported evidence that suggests the “brusque” appearance of imprinting in mammals coinciding with the appearance of viviparity in the Jurassic; the presence of a placental interface suddenly provided an opportunity for paternally-derived genes in the offspring to manipulate transfer of maternal resources, providing enhanced nourishment to the offspring of a particular male to the detriment of the mother and any future offspring by other males (Übeda and Wilkins, 2008). Thus, epigenetic inheritance mechanisms were favored because they could mitigate the severity of this conflict (Day and Bonduriansky, 2004). Given the rarity of lifetime monogamy among mammals, the vast majority demonstrate the conditions polyandrous mating and conflict between the paternal and maternal genomes, which led to imprinting (Haig and Graham, 2002).

The potential role of biparental care and paternal investment in offspring and the evolution of imprinting is a matter of debate (Horsthemke, 2009). Paternal care is considered rare in mammals overall, as it has been observed in less than 10% of all genera. However, levels approaching 40% have been reported in several orders, including Primates, Carnivora, and Rodentia (Kleiman and Malcolm, 1981). Übeda (2008) expanded on the kinship theory of Moore and Haig, describing a model in which the evolution of genomic imprinting occurs even in the presence of monogamous mating. When fathers participate in caring for offspring, genes governing resource allocation from mother to offspring would be expected to be paternally expressed and favor enhanced growth of offspring during gestation and lactation, stages in which the father can play only an indirect role in offspring growth by providing food to the mother. Following weaning, when fathers are able to provide solid food directly to the offspring
in whom they will now invest time and care, reversion of expression patterns is expected at imprinted loci in order to prevent overgrowth of the young.

This conjecture is supported by the observation that murine and human GRB10, which is maternally expressed in the trophoblast during pregnancy and acts to restrict fetal growth, switches to paternal expression as the young begin to mature (Monk et al., 2009). Maternally-expressed genes have also been shown to favor earlier weaning and sexual maturation, likely to improve the mother’s fecundity by reducing the time between pregnancies at the cost of continued maternal nutrient flow to her offspring (Haig, 2010). Moore and Mills (1999) commented on the arguments in the literature over whether polyandrous mating and lack of biparental care are necessary for the evolution of genomic imprinting by pointing out that these social behaviors may have evolved following the occurrence of genomic imprinting. Given the many pathologies associated with LOI, a switch to biallelic expression in taxa now demonstrating monogamous pairing and biparental care may not necessarily be favored at imprinted loci.

Though the reasons genomic imprinting remains an active mechanism in mammalian genomes are under intense debate, many researchers agree that imprinting has likely introduced a number of reproductive incompatibilities which have had an influential effect on speciation in mammals (Jirtle et al., 2000). Jablonka and Lamb (1998) have proposed that the accumulation of different imprints between two parts of a population that have become geographically separated will lead to a reduction in reproductive compatibility and hybrid zygotes which may fail to survive, further diverging the two groups.
Zeh and Zeh (2001) argue that under the parental conflict hypothesis, a coevolutionary arms race would result; parental genomes are continually vying over the levels of resource extraction from the mother, and the level of compatibility between the maternal and paternal genomes may depend on specific breeding pairs. The suppressive activity of maternal genomes is likely controlled by polygenic traits, and paternal genomes will similarly vary in their ability to “fight” maternal genomes for increased resource allocation to the offspring. This presents the intriguing possibility that mating between individuals on opposite ends of the spectrum—for example, highly demanding paternal genomes and poorly competitive maternal genomes—could result in over-demanding progeny and overtaxed maternal resources, rendering the fetus inviable and further contributing to reproductive incompatibility.

1.6 Molecular Evolution of Imprinted Loci

On a molecular level, the evolutionary process can be described as changes in DNA nucleotides over time. Following divergence between two sequences, the number of nucleotide substitutions that occur in each sequence can be determined in order to estimate rates of molecular evolution, allowing the evolutionary history of organisms to be reconstructed (Li and Graur, 1991). Substitutions between the two pyrimidine nucleotides or two purine nucleotides are termed transitions, while substitutions that replace a pyrimidine with a purine, or vice versa, are termed transversions.

Transitional changes have been found to occur more frequently than transversional changes, a bias that is thought to be due to the chemical properties of complementary
base pairing (Rosenberg et al., 2003; Yang, 2006). Watson and Crick (1953) were the first to propose that point mutations might demonstrate a higher frequency of transitions because the tautomeric forms of the four bases required for transversional changes were strongly unfavored. Fresco and Topal (1976) elaborated on this model by stating that the lower frequency of transversions probably results from the need to first form imino or enol tautomers and then to rotate the base in relation to the attached deoxyribose. Transition bias has been documented in nuclear, mitochondrial, and chloroplast sequences across so many species, including prokaryotes, eukaryotes, and viruses, that it is now recognized as a general property of DNA evolution (Wakeley, 1996).

Changes in protein-coding nucleotide sequence can occur either synonymously or nonsynonymously. Also termed silent substitutions, synonymous substitutions do not change the amino acid encoded by the region and are only weakly constrained by natural selection (Tang and Wu, 2006). Nonsynonymous substitutions, also termed replacement substitutions, do result in an amino acid change. Rates of nonsynonymous substitution are known to be highly variable across the genome. Nonsynonymous substitution rate is effectively zero in the strongly conserved histone proteins, whereas interferon-γ demonstrates as many as $2.79 \times 10^{-9}$ substitutions per nonsynonymous site per year (Li and Graur, 1991). Synonymous substitution rates can be highly variable, though not to the extent of nonsynonymous rates, and they have also been shown to undergo stochastic fluctuation due to their weak differences in fitness as compared to wild-type alleles (de la Chaux et al., 2007).

Rates of substitution are heavily influenced by the nucleotide content of the sequence in question, a model termed neighbor dependence. Researchers have known for
some time that a positive correlation exists between the presence of CpG dinucleotides in the genome and increased susceptibility—up to tenfold—to substitution (Arndt et al., 2003). Spontaneous deamination of methylated CpG leads to TpG dinucleotides, or CpA on the reverse strand (Arndt and Hwa, 2005), a process described by Rosenberg et al. (2003) as “fundamentally different” than the far less frequent mutational events taking place in other regions of the genome. Given the widely recognized hypermutability of CpG dinucleotides, as well as the fact that single-nucleotide mutations comprise the majority of sequence lesions in germ lines that lead to human genetic disease, nucleotide substitution in imprinted genes may have powerful effects that extend beyond an influential role in mammalian speciation (Krawczak et al., 1998).

1.7 Medical and Environmental Considerations

LOI has been found to play a significant role in human genetic disease. LOI of Igf2 is associated nearly uniformly with the development of Wilms’ tumor, the most common solid tumor during childhood (Bjornsson et al., 2007). LOI at a variety of other loci is implicated in many adult cancers, including hepatoblastoma, uterine, cervical, esophageal, prostate, lung, and colorectal (Cheng et al., 2010; Feinberg, 2002). Global LOI engineered through eliminating Dnmt1 expression has been shown to result in higher rates of cellular proliferation, shortened cell cycle times, and cellular immortality (Holm et al., 2005). It is now widely accepted that deregulation of imprinting at critical loci, e.g., silencing of tumor-suppressor genes as a result of aberrant hypermethylation, does contribute to tumorigenesis (Jones and Baylin, 2007). Experts have also suggested
that accumulation of epigenetic alterations due to repeated regulatory errors over the
course of an organism’s lifetime may account for the observed increase in cancer risk
correlated with increased age (Feinberg, 2004; Karlic and Varga, 2010).

Prader-Willi, Beckwith-Wiedemann, and Angelman syndromes (PWS, BWS, AS) have all been shown to correlate with loss of imprinted gene expression
(Eggerman et al., 2008; Horsthemke and Wagstaff, 2008). PWS and AS occur in
1/15,000 births and are associated with severe developmental and behavioral deficits;
defects in imprinted inheritance at the long arm (q) of chromosome 15 (q11 - q13) have
been widely identified in PWS and AS patients (Nicholls et al., 1998). BWS occurs with
a frequency of 1/13,700 births and is associated with LOI at the short arm (p) of
chromosome 11 (p15.5), leading to overgrowth, abdominal wall defects, and an increased
incidence of embryonal tumors (Sparago et al., 2007). Hypomethylation of the
paternally-derived H19 DMR has been linked to the development of Silver-Russell
syndrome (SRS), a congenital developmental disorder (Yamazawa et al., 2008). The
incidence of SRS and BWS appears to be increased in pregnancies resulting from in vitro
fertilization (IVF) and other assisted reproductive technologies (ART) in which egg and
sperm are manipulated in the laboratory (Gicquel et al., 2003; Kagami et al., 2007;
Niemitz and Feinberg, 2004).

In recent years, with vastly increased amounts of data available, researchers have
turned their efforts toward characterizing the interplay between DNA methylation,
histone modifications, and the regulation of gene expression patterns from embryogenesis
through adulthood (Dolinoy et al., 2007). Launched in 2006, the Human Epigenome
Project seeks to catalog and describe the methylation patterns of all major human tissues
(Brena et al., 2006). It is hoped that the data obtained from this endeavor will aid researchers in achieving a more advanced understanding and characterization of the vast majority of human diseases, which cannot be explained by variation in a single gene. In the years preceding the Human Epigenome Project, it was becoming increasingly clear to experts that even in cases where a single nucleotide polymorphism (SNP) could be identified, its ability to predict a clinically significant phenotype was generally very poor (Chanock and Wacholder 2002).

Moving beyond the traditional model of genetic disease, susceptibility to aging and illness is now considered the complex result of alterations in methylation state and gene regulation, which are in turn thought to be heavily influenced by an organism’s nutrition, environment, and toxicological exposures (Baccarelli and Bollati, 2009; Esteller, 2006; Jirtle and Skinner, 2007; Kaati, 2010). Deviations from the healthy epigenetic state are strongly associated not only with tumorigenesis but have also been suggested to play a role in the pathogenesis of asthma, autism, schizophrenia, bipolar disorder, autoimmune disorders, metabolic syndrome, and many other conditions thought to result from “triggers” imposed by an individual’s environment and chemical exposures (Cao et al., 2007; Liu et al., 2008; Miller and Ho, 2008; Oh and Petronis, 2008; Vliet et al., 2007). Researchers in the emerging fields of environmental epigenetics and toxicogenomics hope that a comprehensive model of these gene-by-environment interactions will improve our understanding of complex human illness (Reamon-Buettner et al., 2008).
2.1 Bisulfite Sequencing of Imprinted Loci

Due to the overwhelming evidence in favor of methylation state as the regulator of imprinting in *Peromyscus*, as well as reports of widespread LOI in *Peromyscus* during hybridization (see Parent-of-Origin Effects and Hybrid Dysgenesis, Chapter 1.4), recent studies have sought to examine the methylation state of many of the same genes in *Mus* hybrids. The work of Shi *et al.* (2005) has shed light on hybrid dysgenesis and the underlying LOI in laboratory crosses between the species *M. m. musculus* and *M. spretus*. While these data provide fascinating implications for population genetic studies of *M. m. musculus* and *M. spretus*, *M. spretus* individuals are often considered undesirable experimental animals for biomedical work due to their aggressive temperament, poor breeding rates in captivity, and increased sensitivity to stress (Dejager *et al.*, 2009; Palomo *et al.*, 2009).

Two *M. musculus* subspecies strains, *M. m. musculus* and *M. m. domesticus*, are far more commonly used in laboratory genetics research (Wade and Daly, 2005). Genomic DNA (gDNA) samples were obtained from the hybrid offspring of *M. m. musculus* and *M. m. domesticus* with the objective of examining methylation of CpG dinucleotides at the DMRs of multiple imprinted loci. Bisulfite sequencing was performed in order to determine whether the normal methylation patterns of the parental strains are lost as a result of intersubspecific hybridization. Genes investigated included *Igf2r, H19, Rasgrf1*, and *Snrpn*. 
Bisulfite sequencing employs sodium bisulfite (NaHSO$_3$) treatment to force the conversion of unmethylated cytosine nucleotides to uracil through deamination, and the converted DNA is then sequenced. At imprinted loci, DNA methyltransferases add a methyl group (CH$_3$) donated by S-adenosyl-L-methionine to the fifth carbon of cytosine in CpG dinucleotides, forming the so-called fifth base, 5-methylcytosine (Ho and Tang, 2007). At unmethylated cytosines, sodium bisulfite reacts readily with the 5,6-double bond (Raizis et al., 1995). The presence of the methyl group in 5-methylcytosine prevents deamination by sodium bisulfite, effectively protecting the cytosine against conversion (see Figure 3).

![Figure 3: Bisulfite-induced deamination of cytosine to uracil.](image)

At CpG dinucleotides, DNA methyltransferases catalyze the addition of methyl groups to the fifth carbon of cytosine, leading to the formation of 5-methylcytosine and transcriptional silencing of promoters. Sodium bisulfite treatment affects only unmethylated cytosines, resulting in conversion to uracil.
Following bisulfite treatment of gDNA, methylated cytosine is treated differently than uracil—formerly unmethylated cytosine nucleotides—by the DNA polymerase in the polymerase chain reaction (PCR). While methylated cytosine is still paired with guanine, uracil is paired with adenine and will thereafter be amplified as thymine. Direct nucleotide sequencing of the resulting PCR fragments provides methylation status.

2.1.1 H19

H19 is a paternally-imprinted, non-coding locus on the distal portion of mouse chromosome 7. It lies approximately 90 kb from Igf2, Ins2, and Mash2, which are also imprinted. The H19 DMR (GenBank Accession No. U19619), located 2-4 kb upstream from the transcription start site, has been implicated as an ICR influencing not only H19 transcription but also methylation and transcription at Igf2 through association with CTCF (Davis et al., 2000; Thorvaldsen et al., 1998). H19 and Igf2 together constitute an imprinted domain and are reciprocally expressed in both mice and humans (Zemel et al., 1992). The H19 DMR holds the distinction of being the most ancient case of genomic imprinting known: while it is conserved in the therians, all other imprinting clusters have been observed only in eutherians (Smits et al., 2008). H19 mRNA provides the substrate for cleavage into a regulatory miRNA, miR-675, whose function is not yet understood (Cai and Cullen, 2007; Peters and Robson, 2008).

2.1.2 Igf2r

Igf2r encodes the cation-independent mannose-6-phosphate receptor, expressed ubiquitously in the human and mouse (Brown et al., 2008; Zwart et al., 2001). Two
DMRs have been identified: the first, located upstream of the promoter region, is unique to the mouse genome; the second (GenBank Accession No. L06446), spanning portions of exon 2 and intron 2, is conserved in both mice and humans (Lucifero et al., 2002; Smrzka et al., 1995; Turner et al., 2010).

In mice, Igf2r is paternally imprinted (Lerchner and Barlow, 1997). Conflicting evidence from studies of IGF2R allele-specific expression in humans suggest a functional polymorphism in the ability to imprint IGF2R earlier in development, with some placental and fetal tissues repressing the paternal allele only partially, while in others imprinting is completely absent during all stages (Reik and Walter, 2001; Xu et al., 1993). Some researchers have suggested IGF2R imprinting diminished throughout human evolution because the majority of human pregnancies result in the birth of only one child (a singleton), whereas mice bear litters and have experienced increased pressure to maintain imprinting (Monk et al., 2006; Wood and Oakey, 2006).

Igf2r is responsible for binding mannose-6-phosphate and transporting the molecules to lysosomes, both intracellularly and at the cell surface (Dahms et al., 1989; Wylie et al., 2003). MacDonald et al. (1988) reported that the mannose 6-phosphate receptor is multifunctional; it contains a second binding site, which binds and internalizes IGF-II. Downregulation of Igf2r expression, and consequently high circulating levels of IGF-II, are correlated with the growth of tumors in both humans and mice, consistent with its role as a growth suppressor and the role of IGF-II as a growth promoter (Hassan and Howell, 2000; Toretsky and Hellman, 1996).
2.1.3 Rasgrf1

In neonatal mice, Rasgrf1, also referred to as Grf1, is maternally imprinted. By adulthood, expression in some tissues, including the brain, becomes biallelic (Drake et al., 2009). The protein product RasGRF1 acts as a guanine nucleotide exchange factor and is responsible for activating the G proteins Ras and Rac by promoting the GTP-bound state (Overbeck et al., 1995; Yang and Mattingly, 2006). Mice suffering from disrupted expression of Rasgrf1 demonstrate learning and memory deficits, indicating that Rasgrf1 protein may play a crucial role in synaptic remodeling (Brambilla et al., 1997).

The Rasgrf1 DMR (GenBank Accession No. NT_038476) and its associated tandem repeats have been heavily studied as a model of how nearby sequences are capable of influencing whether methylation is properly established in CGIs of the DMR (Delaval and Feil, 2004; Dockery et al., 2009). The Rasgrf1 DMR, on mouse chromosome 9, begins with CGIs approximately 30 kb upstream of the promoter; a second set of CGIs is found following the tandem repeats, just upstream of the promoter region. Adjacent to the 3’ end of the DMR is a 41-mer sequence tandemly repeated 40 times which has been found to be capable of inducing methylation of the paternal Rasgrf1 DMR, possibly by recruiting transcription factors (Herman et al., 2003; Yoon et al., 2002).

2.1.4 Snrpn

Snrpn is maternally imprinted in nearly all fetal and adult tissues, both in mice and in humans (Cattanach et al., 1992; Glenn et al., 1993). Snrpn encodes the 28-kDa
protein SmB, which is involved in forming the pre-mRNA splicing complex (Grimaldi et al., 1993). Deletions of the Snrpn DMRs in humans, found on chromosome 15, are associated with the development of PWS and AS. The first DMR extends from the promoter region to exon 1 (GenBank Accession No. AF081460 (Hiura et al., 2006). It has been found to be partially deleted in multiple PWS families who demonstrate abnormal silencing of the paternal allele. Families with AS have been found to carry deletions of the maternal chromosome encompassing a second DMR approximately 35 kb upstream of the promoter (Ohta et al., 1999).

Like the Rasgrf1 DMR, the establishment of proper methylation patterns at the Snrpn DMRs has been linked to an upstream region (Reinhart et al., 2006). Delaval and Feil (2004) have suggested that repeat or flanking sequences may recruit trans-acting factors in one of the germ lines only and are therefore responsible for inducing local methylation and establishing the initial germline imprint. More recent bioinformatic analyses by Hutter et al. (2010) have indicated that tandem repeats are themselves significantly enriched in CGIs in approximately one-third of human and mouse imprinted genes. Whether these CGIs enable repeat sequences to influence methylation of nearby DMRs remains under debate.

2.2 Methods and Materials

Liver samples from male and female M. m. musculus, M. m. domesticus, and M. m. musculus/M. m. domesticus F1 hybrids were obtained from the laboratory of Dr. Bret Payseur of the University of Wisconsin-Madison. M. m. domesticus individuals were of the Watkins Star B strain out of Centreville, Maryland (Scott and Potter, 1984).
*M. m. musculus* individuals were of the PWD strain, originating near Prague, Czech Republic (Gregorová and Forejt, 2000). Extraction of gDNA was performed using the DNeasy Blood & Tissue Kit (Qiagen, Valencia, CA).

In order to examine methylation state, the EZ DNA Methylation Kit (Zymo Research, Orange, CA) was used to convert unmethylated cytosine to uracil within extracted gDNA samples. PCR amplification was performed to assay multiple CpG sites within each DMR (see Appendix A for primer sequences and conditions). All samples were obtained in triplicate. CGIs within each DMR were identified using Methyl Primer Express 1.0 (Applied Biosystems, Foster City, CA). Primers were specific to the sequence of the mutagenized (post-bisulfite treatment) DNA. Negative controls were obtained by amplifying the same region from unconverted gDNA. Primers for negative control sequences were designed with Primer3 (Rozen and Skaletsky, 2000).

Amplification of bisulfite-treated DNA was performed with ZymoTaq PreMix (Zymo Research, Orange, CA), a hot start Taq polymerase designed specifically to reduce non-specific product when amplifying bisulfite-treated DNA. Negative controls were amplified with Platinum Taq (Invitrogen, Carlsbad, CA). Direct nucleotide sequencing was performed with BigDye Terminator v3.1 (Applied Biosystems, Foster City, CA) on an ABI 3100-Avant sequencer. Methylated mouse gDNA controls were obtained through Zymo Research (Orange, CA). See Appendix A for confirmation of lack of conversion with methylated gDNA control (Figure 14). Sequencing of all samples was performed in triplicate. Forward and reverse sequences were compared and manually edited to produce the final results.
2.3 Results

Following nucleotide sequencing, the results of bisulfite-treated, converted gDNA and untreated protocols were compared to determine whether the methylation state of cytosine nucleotides at the loci of Igf2r, H19, Rasgrf1, and Snrpn was disrupted in hybrids. All negative control sequences were compared against the NCBI non-redundant nucleotide database with the Basic Local Alignment Search Tool (BLAST) to verify the genomic location of amplified regions.

2.3.1 H19

Top hits from BLAST indicated that the DMR of the M. m. musculus H19 gene was successfully amplified and sequenced (GenBank Accession Numbers U19619.1, AY849916.1). Sequences from M. m. musculus, M. m. domesticus, and their F1 hybrids were aligned and investigated for differences in the retention of CpG sites, which indicate that methylation of the cytosine protected it from conversion to uracil and, subsequently, detection as thymine during sequencing. See Figures 4A-B for full sequencing results from the H19 DMR and comparison of CpG sites in the parental strains and hybrid offspring. A total of nine CpG sites were assayed. Shaded CpGs indicate successful detection of at least one methylated allele during sequencing.
**Figure 4A: Comparison of CpG methylation of the H19 DMR between a M. m. musculus female, M. m. domesticus male, and their hybrid offspring.** Shaded (gray) areas mark CpGs retained through bisulfite treatment, indicating the presence of 5-methylcytosine. TpG sites (highlighted in yellow and blocked) indicate loss of cytosine methylation and resulting deamination in the hybrid offspring. N indicates failure to call the base. W indicates ambiguity between A and T.
H19 M. m. domesticus female

TTTGTTAGGATTTAATGTGTTTTTTGAGTTTTTTTTGGAATTAGTGTGGGTTTTATACGGGGGAGT
TGCCTGCTGGTCTGGTAGTTAATACGATTCGCTTTAATTTAAAGAGTTTTTTTATTTTTGGAAT
TTATAATGTTAACGTGTGCGTTATTATTTAGTTATTTAGGAGGGTTATAAATGTTATTAGGGGGGT
AGGATATTTTTTTTATAGG

H19 M. m. musculus male

TTTGTTAGGATTTAATGTTTTTTATTTTTGGAATTAGTGTGGGTTTTATACGGGGGAGT
TGCCTGCTGGTCTGGTAGTTAATACGATTCGCTTTAATTTAAAGAGTTTTTTTATTTTTGGAAT
TTATAATGTTAACGTGTGCGTTATTATTTAGTTATTTAGGAGGGTTATAAATGTTATTAGGGGGGT
AGGATATTTTTTTTATAGG

H19 F₁ female (M. m. domesticus ♀ x M. m. musculus ♂)

TTTGTTNNGATTTAATGTGTTTTTTGAGTTTTTTTTGGAATTAGTGTGGGTTTTATACGGGGGAGT
TGCCTGCCCTGGTCTGGTAGTTAATACGATTCGCTTTAATTTAAAGAGTTTTTTTATTTTTGGAAT
TTATAATGTTAACGTGTGCGTTATTATTTAGTTATTTAGGAGGGTTATAAATGTTATTAGGGGGGT
AGGATATTTTTTTTATAGG

H19 F₁ male (M. m. domesticus ♀ x M. m. musculus ♂)

TTTGTTAGGATTTAATGTGTTTTTTGAGTTTTTTTTGGAATTAGTGTGGGTTTTATACGGGGGAGT
TGCCTGCCCTGGTCTGGTAGTTAATACGATTCGCTTTAATTTAAAGAGTTTTTTTATTTTTGGAAT
TTATAATGTTAACGTGTGCGTTATTATTTAGTTATTTAGGAGGGTTATAAATGTTATTAGGGGGGT
AGGATATTTTTTTTATAGG

Figure 4B: Comparison of CpG methylation of the H19 DMR between a M. m. domesticus female, M. m. musculus male, and their hybrid offspring. Shaded (gray) areas mark CpGs retained through bisulfite treatment, indicating the presence of 5-methylcytosine. TpG sites (highlighted in yellow and blocked) indicate loss of cytosine methylation and resulting deamination in the hybrid offspring. N indicates failure to call the base. W indicates ambiguity between A and T.
Heterozygosity was resolved during sequencing by the detection of fluorescence from multiple bases at a single site. *M. m. musculus* and *M. m. domesticus* individuals demonstrated methylation of one allele for all nine CpGs, resulting in the simultaneous detection of cytosine and thymine (see Figure 5, below). Loss of CpG methylation was determined based on the complete absence of cytosine fluorescence and was detected at three CpGs in each of the *M. m. musculus ♀ x M. m. domesticus ♂* offspring and four and five CpGs in the male and female offspring, respectively, of the *M. m. domesticus ♀ x M. m. musculus ♂* cross.

*Figure 5: Detection of CpG methylation at one allele following bisulfite sequencing.* The presence of fluorescence from both cytosine (blue) and thymine (red) simultaneously, shown in the box above, indicates methylation at the CpG of one allele. Following bisulfite conversion, 5-methylcytosine was successfully detected from the methylated allele, while the unmethylated cytosine of the other allele was converted to uracil, amplified as adenine during PCR, and finally detected as thymine during sequencing. The sample shown is a portion of the *H19* DMR (sixth CpG site) in a *M. m. domesticus* female.
2.3.2 *Igf2r*

Nucleotide BLAST was also performed with *Igf2r* negative control sequences. Top hits indicated that the second intron of *Igf2r* was successfully sequenced (GenBank Accession Number AF151173). Within the *Igf2r* intronic DMR, the seven CpG sites previously described by Lucifero *et al.* (2002) were assayed for methylation state. Methylation was detected in all samples, both parental strains and F₁ hybrids, at all CpG sites. No evidence was found to indicate that CpG methylation was lost in the F₁ hybrids.

An eighth CpG site at *Igf2r* was noted upstream of the seven CpGs previously described and occurs as the first CpG site shown in Figure 6. All samples contained this CpG site, and all demonstrated evidence that methylation was retained through bisulfite treatment. This CpG has been boxed in Figure 6. *Igf2r* DMR primers were previously described by Lucifero *et al.* (2002), who employed nested PCR and reported their results for sequences beginning approximately 26 bp downstream of this first CpG, consistent with the fragment that would be amplified using the inside forward primer.

A larger fragment (corresponding to the outside forward primer) was successfully amplified here, for a total of 208 bp rather than 158 bp, and no ambiguous sites were found in the first 50 bp of any of these samples. Furthermore, the first 50 bp are still considered to lie within the intronic DMR. The eighth CpG site has therefore been reported here.
Igf2r M. m. musculus female

TTAATGAGGAATTTTTAATTTTGGAGGAGTGTGGTATTTTTATGTATAGTTAGGATAGCGTTAAATTTTG
TTTTTTTTTTTTTTTGTAAACCGCGTATTTTTGGATGTTTTTTTTTTTTTTGAACGGTGATTTTTTGAGTTT
ATTTTTTTTGTAAATATGGTATTTTTGAGCGTTGTTTTGATTATAGAATTTTCCGAATTTTTTTTT

Igf2r M. m. domesticus male

TTAATGAGGAATTTTTAATTTTGGAGGAGTGTGGTATTTTTATGTATAGTTAGGATAGCGTTAAATTTTG
TTTTTTTTTTTTTTTGTAAACCGCGTATTTTTGGATGTTTTTTTTTTTTTTGAACGGTGATTTTTTGAGTTT
ATTTTTTTTGTAAATATGGTATTTTTGAGCGTTGTTTTGATTATAGAATTTTCCGAATTTTTTTTT

Igf2r F₁ female (M. m. musculus ♀ x M. m. domesticus ♂)

TTAATGAGGAATTTTTAATTTTGGAGGAGTGTGGTATTTTTATGTATAGTTAGGATAGCGTTAAATTTTG
TTTTTTTTTTTTTTTGTAAACCGCGTATTTTTGGATGTTTTTTTTTTTTTTGAACGGTGAGTTTTTTTT
ATTTTTTTTGTTGATTATAGAATTTTCCGAATTTTTTTTT

Igf2r F₁ male (M. m. musculus ♀ x M. m. domesticus ♂)

TTAATGAGGAATTTTTAATTTTGGAGGAGTGTGGTATTTTTATGTATAGTTAGGATAGCGTTAAATTTTG
TTTTTTTTTTTTTTTGTAAACCGCGTATTTTTGGATGTTTTTTTTTTTTTTGAACGGTGGAATTTTTGAGTTT
ATTTTTTTTGTAAATATGGTATTTTTGAGCGTTGNTTTGATTATAGAATTTTCCGAATTTTTTTTT

Figure 6A: Comparison of CpG methylation of the Igf2r DMR between a M. m. musculus female, M. m. domesticus male, and their hybrid offspring. Shaded (gray) areas mark CpGs retained through bisulfite treatment, indicating the presence of 5-methylcytosine. The first CpG, which occurs in all samples and was not described by Lucifero et al. (2002), who assayed the remaining seven sites, has been boxed. No evidence was found to support loss of methylation in these F₁ hybrids as compared to the methylation state of the parents. N indicates failure to call the base. W indicates ambiguity between A and T.
**Igf2r M. m. domesticus female**

TTAAATGAGGAATTTTTAATTTCGAGGAGTGTGGTATTTTTATGTATAGTTAGGATACGCTTTAAATTTTG
TTTTTTTTTTCTTTTTTGTAACCGCGGTATTTTTGAGTTTTTTTTGTAACCGTGGTATAATTTTGTAGTT
WTTTTTTGTAAAWATGGTATTTTTGACCGTTGGTCTGATTATAGAATTTTTCCGAATTTTTTTTT

**Igf2r M. m. musculus male**

TTAAATGAGGAATTTTTAATTTTCGAGGAGTGTGGTATTTTTATGTATAGTTAGGATACGCTTTAAATTTTG
TTTTTTTTTTCTTTTTTGTAACCGCGGTATTTTTGAGTTTTTTTTGTAACCGTGGTATAATTTTGTAGTT
ATTTTTTTGTAAATATGGAATTTTTGACCGTTGATTATANAATTTTTCCGAATTTTTTTTT

**Igf2r F1 female (M. m. domesticus ♀ x M. m. musculus ♂)**

TTAAATGAGGAATTTTTAATTTCGAGGAGTGTGGTATTTTTATGTATAGTTAGGATACGCTTTAAATTTTG
TTTTTTTTTTCTTTTTTGTAACCGCGGTATTTTTGAGTTTTTTTTGTAACCGTGGTATAATTTTGTAGTT
ATTTTTTTGTAAATATGGAATTTTTGACCGTTGGTCTGATTATAGAATTTTTCCGAATTTTTTTTT

**Igf2r F1 male (M. m. domesticus ♀ x M. m. musculus ♂)**

TTAAATGAGGAATTTTTAATTTCGAGGAGTGTGGTATTTTTATGTATAGTTAGGATACGCTTTAAATTTTG
TTTTTTTTTTCTTTTTTGTAACCGCGGTATTTTTGAGTTTTTTTTGTAACCGTGGTATAATTTTGTAGTT
ATTTTTTTGTAAATAMGGTATTTTTGACCGTTGGTCTGATTATAGAATTTTTCCGAATTTTTTTTT

**Figure 6B: Comparison of CpG methylation of the Igf2r DMR between a M. m. domesticus female, M. m. musculus male, and their hybrid offspring.** Shaded (gray) areas mark CpGs retained through bisulfite treatment, indicating the presence of 5-methylcytosine. The first CpG, which occurs in all samples and was not described by Lucifero et al. (2002), who assayed the remaining seven sites, has been boxed. No evidence was found to support loss of methylation in these F1 hybrids as compared to the methylation state of the parents. N indicates failure to call the base. W indicates ambiguity between A and T.
2.3.3 Rasgrf1

Primer sequences previously described by Dockery et al. (2009) were used to assay 12 CpG sites within the Rasgrf1 DMR region located immediately 5’ to the promoter sequence. The genomic region amplified in Rasgrf1 negative control sequences was verified using BLAST. The top hits were found to correspond to the sequence uploaded to GenBank by Plass et al. (1996) (U55232.1), who first reported Rasgrf1 as imprinted, and to M. m. musculus chromosome 9 (AC102545.9), consistent with the known location of Rasgrf1. The Plass et al. (1996) sequence was verified to be contained within the genomic location reported by Dockery et al. (2009) (NT_038476), beginning with nucleotide position 9,476,349.

Alignments of Rasgrf1 bisulfite-converted sequences can be seen in Figure 7. Both male and female F_1 offspring of the M. m. musculus ♀ x M. m. domesticus ♂ cross demonstrate loss of methylation at the first CpG site, while the F_1 offspring of the reciprocal cross (M. m. domesticus ♀ x M. m. musculus ♂) demonstrate loss of methylation at the first and fourth CpG sites.
Rasgrf1 *M. m. musculus* female

TTTTTTTTTTTTTTTACCCGGGGAAGGATCTGGGAGTTGGGATTAAATGGGAG
GGTATTTAGGCTGGTAGAGGAATAGGAGTCCCCGGGATGTTAATTTAGGGATCGTTTTTTG
TAGGGTAATTTTTCTAGGAGTGGCCGGGGGG

Rasgrf1 *M. m. domesticus* male

TTTTTTTTTTTTTTTACCCGGGGAAGGATCTGGGAGTTGGGATTAAATGGGAG
GGTATTTAGGCTGGTAGAGGAATAGGAGTCCCCGGGATGTTAATTTAGGGATCGTTTTTTG
TAGGGTAATTTTTCTAGGAGTGGCCGGGGGG

Rasgrf1 F₁, female (*M. m. musculus* ♀ x *M. m. domesticus* ♂)

TTTTTTTTTTTTTTTACCCGGGGAAGGATCTGGGAGTTGGGATTAAATGGGAG
GGTATTTAGGCTGGTAGAGGAATAGGAGTCCCCGGGATGTTAATTTAGGGATCGTTTTTTG
TAGGGTAATTTTTCTAGGAGTGGCCGGGGGG

Rasgrf1 F₁, male (*M. m. musculus* ♀ x *M. m. domesticus* ♂)

TTTTTTTTTTTTTTTACCCGGGGAAGGATCTGGGAGTTGGGATTAAATGGGAG
GGTATTTAGGCTGGTAGAGGAATAGGAGTCCCCGGGATGTTAATTTAGGGATCGTTTTTTG
TAGGGTAATTTTTCTAGGAGTGGCCGGGGGG

Figure 7A: Comparison of CpG methylation of the Rasgrf1 DMR between a *M. m. musculus* female, *M. m. domesticus* male, and their hybrid offspring. Shaded (gray) areas mark CpGs retained through bisulfite treatment, indicating the presence of 5-methylcytosine. TpG sites (highlighted in yellow and blocked) indicate loss of cytosine methylation and resulting deamination in the hybrid offspring. N indicates failure to call the base. W indicates ambiguity between A and T.
Rasgrf1 M. m. domesticus female

TTTTTTTTTTTAAAGATTAGGCGTTTTTAGTATTTTCCCGGGAAGGATCGTTGGATTAATTGGGAG
GTTGATTGTACGGGTAGAGAGGTAGGGATTCGCGGTAGGGCTGTTATTAGTTGCGATCGTTTTTGG
TAGGGTAAATTTTCTAGGAGTGGCGGGGGGG

Rasgrf1 M. m. musculus male

TTTTTTTTTTTAAAGATTAGGCGTTTTTAGTATTTTCCCGGGAAGGATCGTTGGATTAATTGGGAG
GTTGATTGTACGGGTAGAGAGGTAGGGATTCGCGGTAGGGCTGTTATTAGTTGCGATCGTTTTTGG
TAGGGTAAATTTTCTAGGAGTGGCGGGGGGG

Rasgrf1 F1 female (M. m. domesticus ♀ x M. m. musculus ♂)

TTTTTTTTTTAAGAATTAGGCGGATTTTAGTATTTTCCCGGGAAGGATCGTTGGATTAATTGGGAG
GTTGATTGTACGGGTAGAGAGGTAGGGATTCGCGGTAGGGCTGTTATTAGTTGCGATCGTTTTTGG
TAGGGTAAATTTTCTAGGAGTGGCGGGGGGG

Rasgrf1 F1 male (M. m. domesticus ♀ x M. m. musculus ♂)

TTTTTTTTTTAAGAATTAGGCGGATTTTAGTATTTTCCCGGGAAGGATCGTTGGATTAATTGGGAG
GTTGATTGTACGGGTAGAGAGGTAGGGATTCGCGGTAGGGCTGTTATTAGTTGCGATCGTTTTTGG
TAGGGTAAATTTTCTAGGAGTGGCGGGGGGG

Figure 7B: Comparison of CpG methylation of the Rasgrf1 DMR between a
M. m. domesticus female, M. m. musculus male, and their hybrid offspring. Shaded
(gray) areas mark CpGs retained through bisulfite treatment, indicating the presence of 5-
methylcytosine. TpG sites (highlighted in yellow and blocked) indicate loss of cytosine
methylation and resulting deamination in the hybrid offspring. N indicates failure to call
the base. W indicates ambiguity between A and T.
2.3.4 *Snrpn*

The final DMR investigated was that of *Snrpn*, which begins upstream of the promoter and extends toward the first exon. Sequences are shown in Figure 8. The top hit when performing nucleotide BLAST confirmed that the sequence amplified includes the *Snrpn* DMR sequence upstream of the promoter region and first exon (AF332579.1).

While *M. m. musculus ♀ x M. m. domesticus ♂* F₁ offspring demonstrated no loss of CpG methylation as compared to *M. m. musculus* and *M. m. domesticus* parents (Figure 9A), the male offspring of the reciprocal cross (*M. m. domesticus ♀ x M. m. musculus ♂*) showed no evidence of the presence of cytosine at the third CpG site in the parents and appears to have lost methylation (Figure 8B). The female hybrid offspring of the same cross, however, demonstrated clear evidence of cytosine fluorescence and has therefore retained methylation.
**Snrpn** *M. m. musculus* female

TTTTTTTTATATTTGGAGATTAAAAATTTTTTTTTTTTTTTTATATAGTAATAAATTTGTGATGTGGTATAATT
ATTTGGAAATAATTTTTTTTTAATTTTTAGTATTGGTAAATAGTAATATATTATATTATATAGTG
ATAATATTTTTTTTTTTTTAAAAATTACGGTTAAAATTTTGTAGAGTTTTTTTTGTTAGTT
GTTTTTTGTTAGGAATTTCCGTTAAAAGGGATATAGATTTTTGTATTTGCCGTAATAAATGTCGGATATGTGAG
TTATTTTGGGACTATCCGTTAGGAGTCCGGGATAAAT

**Snrpn** *M. m. domesticus* male

TTTTTTTTATATTTGGAGATTAAAAATTTTTTTTTTTTTTTTATATAGTAATAAATTTGTGATGTGGTATAATT
ATTTGGAAATAATTTTTTTTTAATTTTTAGTATTGGTAAATAGTAATATATTATATTATATAGTG
ATAATATTTTTTTTTTTTTAAAAATTACGGTTAAAATTTTGTAGAGTTTTTTTTGTTAGTT
GTTTTTTGTTAGGAATTTCCGTTAAAAGGGATATAGATTTTTGTATTTGCCGTAATAAATGTCGGATATGTGAG
TTATTTTGGGACTATCCGTTAGGAGTCCGGGATAAAT

**Snrpn** *F₁* female (*M. m. musculus* ♀ x *M. m. domesticus* ♂)

TTTTTTTTATATTTGGAGATTAAAAATTTTTTTTTTTTTTTTATATAGTAATAAATTTGTGATGTGGTATAATT
ATTTGGAAATAATTTTTTTTTAATTTTTAGTATTGGTAAATAGTAATATATTATATTATATAGTG
ATAATATTTTTTTTTTTTTAAAAATTACGGTTAAAATTTTGTAGAGTTTTTTTTGTTAGTT
GTTTTTTGTTAGGAATTTCCGTTAAAAGGGATATAGATTTTTGTATTTGCCGTAATAAATGTCGGATATGTGAG
TTATTTTGGGACTATCCGTTAGGAGTCCGGGATAAAT

**Snrpn** *F₁* male (*M. m. musculus* ♀ x *M. m. domesticus* ♂)

TTTTTTTTATATTTGGAGATTAAAAATTTTTTTTTTTTTTTTATATAGTAATAAATTTGTGATGTGGTATAATT
ATTTGGAAATAATTTTTTTTTAATTTTTAGTATTGGTAAATAGTAATATATTATATTATATAGTG
ATAATATTTTTTTTTTTTTAAAAATTACGGTTAAAATTTTGTAGAGTTTTTTTTGTTAGTT
GTTTTTTGTTAGGAATTTCCGTTAAAAGGGATATAGATTTTTGTATTTGCCGTAATAAATGTCGGATATGTGAG
TTATTTTGGGACTATCCGTTAGGAGTCCGGGATAAAT

**Figure 8A:** Comparison of CpG methylation of the Snrpn DMR between a *M. m. musculus* female, *M. m. domesticus* male, and their hybrid offspring. Shaded (gray) areas mark CpGs retained through bisulfite treatment, indicating the presence of 5-methylcytosine. No evidence was found to support loss of methylation in these *F₁* hybrids as compared to the methylation state of the parents. N indicates failure to call the base. W indicates ambiguity between A and T.
**Snrpn M. m. domesticus female**

TTTTTTTATATTGGGATTAATTTTTTTTTTTTTATATAGTAATTGTGGGATGTGTGTAATT
ATTTGGGAATATTTTTTTAAAAATTAATGTTTTTAGTTATATATTTTATTTTATGATTG
ATAGTGATTATTTTTTTTTATATAGCTAAATTTTTCTAGATAGGAATGTTAATTTTTTTGTAGTT
GTTTTTTGCTAGATATTCTGGTTAAAGGGATAGATTTTGGTTTACCGTTAAAAATGTCGTATGTA
GTTATTTTTGGGAGTNGTGGGAGTCGGCCTAAAT

**Snrpn M. m. musculus male**

TTTTTTTTATATTGGGATTAATTTTTTTTTTTTTTATATAGTAATTGTGGGATGTGTGTAATT
ATTTGGGAATATTTTTTTAAAAATTAATGTTTTTAGTTATATATTTTATTTTATGATTG
ATAGTGATTATTTTTTTTTTTATATAGCTAAATTTTTCTAGATAGGAATGTTAATTTTTTTGTAGTT
GTTTTTTGCTAGATATTCTGGTTAAAGGGATAGATTTTGGTTTACCGTTAAAAATGTCGTATGTA
GTTATTTTTGGGAGTNGTGGGAGTCGGCCTAAAT

**Snrpn F; female (M. m. domesticus ♀ x M. m. musculus ♂)**

TTTTTTTTATATTGGGATTAATTTTTTTTTTTTTTATATAGTAATTGTGGGATGTGTGTAATT
ATTTGGGAATATTTTTTTAAAAATTAATGTTTTTAGTTATATATTTTATTTTATGATTG
ATAGTGATTATTTTTTTTTTTATATAGCTAAATTTTTCTAGATAGGAATGTTAATTTTTTTGTAGTT
GTTTTTTGCTAGATATTCTGGTTAAAGGGATAGATTTTGGTTTACCGTTAAAAATGTCGTATGTA
GTTATTTTGGGAGTNGTGGGAGTCGGCCTAAAT

**Snrpn F; male (M. m. domesticus ♀ x M. m. musculus ♂)**

TTTTTTTTATATTGGGATTAATTTTTTTTTTTTTTATATAGTAATTGTGGGATGTGTGTAATT
ATTTGGGAATATTTTTTTAAAAATTAATGTTTTTAGTTATATATTTTATTTTATGATTG
ATAGTGATTATTTTTTTTTTTATATAGCTAAATTTTTCTAGATAGGAATGTTAATTTTTTTGTAGTT
GTTTTTTGCTAGATATTCTGGTTAAAGGGATAGATTTTGGTTTACCGTTAAAAATGTCGTATGTA
GTTATTTTGGGAGTNGTGGGAGTCGGCCTAAAT

**Figure 8B: Comparison of CpG methylation of the Snrpn DMR between a M. m. domesticus female, M. m. musculus male, and their hybrid offspring.** Shaded (gray) areas mark CpGs retained through bisulfite treatment, indicating the presence of 5-methylcytosine. TpG sites (highlighted in yellow and blocked) indicate loss of cytosine methylation and resulting deamination in the hybrid offspring. No evidence was found to support loss of methylation in the female hybrid, though the male demonstrated lack of methylation at the third CpG site. N indicates failure to call the base. W indicates ambiguity between A and T.
See Table 2 (below) for a summary of CpG methylation at all four loci investigated (*H19, Igf2r, Rasgrfl, Snrpn*).

**Table 2: Summary of CpG methylation at four imprinted loci.** Bisulfite sequencing converted unmethylated cytosine to uracil, which was detected as thymine during nucleotide sequencing. All 5-methylcytosines at CpG sites were protected from conversion. F1 hybrids were compared to *M. m. musculus* and *M. m. domesticus* parents to determine whether CpG methylation was lost as a result of hybridization. Check marks indicate methylation. Crosses indicate loss of methylation in the hybrids.

<table>
<thead>
<tr>
<th>Locus</th>
<th>Sample</th>
<th>CpG Site</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>H19</td>
<td><em>M. m. musc. (P)</em></td>
<td></td>
<td>✓</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td></td>
<td><em>M. m. dom. (P)</em></td>
<td></td>
<td>✓</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td></td>
<td>musc. x dom. (F1)</td>
<td></td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>dom. x musc. (F1)</td>
<td></td>
<td>X</td>
<td>✓</td>
<td>✓</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Igf2r</td>
<td><em>M. m. musc. (P)</em></td>
<td></td>
<td>✓</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td></td>
<td><em>M. m. dom. (P)</em></td>
<td></td>
<td>✓</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td></td>
<td>musc. x dom. (F1)</td>
<td></td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td></td>
<td>dom. x musc. (F1)</td>
<td></td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Rasgrfl</td>
<td><em>M. m. musc. (P)</em></td>
<td></td>
<td>✓</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td></td>
<td><em>M. m. dom. (P)</em></td>
<td></td>
<td>✓</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td></td>
<td>musc. x dom. (F1)</td>
<td></td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td></td>
<td>dom. x musc. (F1)</td>
<td></td>
<td>X</td>
<td>✔</td>
<td>✔</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Snrpn</td>
<td><em>M. m. musc. (P)</em></td>
<td></td>
<td>✓</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td></td>
<td><em>M. m. dom. (P)</em></td>
<td></td>
<td>✓</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td></td>
<td>musc. x dom. (F1)</td>
<td></td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td></td>
<td>dom. x musc. (F1)</td>
<td></td>
<td>✔</td>
<td>✔</td>
<td>X</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
</tbody>
</table>
2.4 Discussion

Loss of CpG methylation was identified at three of the four loci investigated 
(H19, Rasgrf1, Snrpn) in three replicates. CpG methylation was found to be fully retained 
in F1 hybrids for only the DMR of Igf2r.

2.4.1 Conservation of CpG Methylation at Igf2r

CpG methylation, as detected by the presence of cytosine fluorescence following 
bisulfite treatment of gDNA and direct nucleotide sequencing, was retained at all eight 
CpG sites in both M. m. musculus ♀ x M. m. domesticus ♂ and M. m. domesticus ♀ x 
M. m. musculus ♂ F1 hybrids at the Igf2r DMR. Thus, the results obtained for these four 
F1 hybrids are not consistent with the possibility of LOI resulting from loss of CpG 
methylation at Igf2r. Whether this is true in all cases of M. m. musculus/M. m. domesticus 
intersubspecific hybridization will require further study, especially given the compelling 
evidence in favor of LOI at Igf2r in other Mus interspecies hybrids. As described 
previously, Shi et al. (2005) found LOI to be widespread in M. m. musculus/M. spretus 
hybrids, including aberrant expression of Igf2r.

I would suggest that future studies encompassing a larger sample size of 
M. m. musculus/M. m. domesticus F1 hybrids are indicated. It may be the case that LOI at 
Igf2r is simply less common in these intersubspecific hybrids as compared to other 
hybridization events in the genus Mus. M. m. musculus and M. spretus were described as 
having incompatibilities in the molecular mechanisms required for the maintenance of 
genomic imprinting in their F1 hybrids, leading to “abnormal reprogramming” after
fertilization (Shi et al., 2005). Given that M. musculus and M. spretus diverged more
distantly—approximately 1.5 million years ago as compared to 370,000 years for the
M. m. musculus/M. m. domesticus divergence (Goios et al., 2007;
Guenet and Bonhomme, 2003)—it is possible that M. m. musculus/M. m. domesticus F₁
hybrids do not experience perturbation of CpG methylation to the same extent due to
greater conservation of these molecular mechanisms. Without more extensive data,
however, I would hesitate to conclude that there is no difficulty establishing correct
genomic imprinting patterns at Igf2r.

2.4.2 Loss of CpG Methylation at H19, Rasgrf1, and Snrpn

The presence of CpG dinucleotides at all sites assayed in the H19, Rasgrf1, and
Snrpn DMRs in the parent lines supports the proper establishment of methylation at the
imprinted (silenced) allele and maintenance of genomic imprinting in both
M. m. musculus and M. m. domesticus. At none of the sites were exclusively TpG
dinucleotides detected. In the F₁ hybrids, however, CpG methylation was lost at multiple
sites, as indicated by the total lack of cytosine fluorescence following bisulfite treatment.
Without the protection afforded by a methyl group, cytosine was converted to uracil at
multiple sites, resulting in the detection of TpG sites during sequencing. It therefore
appears that imprinting has not been fully established in
M. m. musculus/M. m. domesticus F₁ hybrids at these three loci, unlike the results
obtained for Igf2r.
Interestingly, while in these F₁ hybrids the *H19* DMR suffered the most extensive loss of CpG methylation, Shi *et al.* (2005) reported *H19* to be one of three loci whose imprinted expression was *not* found to be disrupted in the tissues of F₁ hybrids of *M. m. musculus* and *M. spretus*. Though my results examined only CpG methylation, not expression from the maternal and paternal alleles, it can be inferred that LOI at all three loci is possible. Investigation of the transcripts produced will be required to determine whether changes in CpG methylation correlate with aberrant expression of both parental alleles in *M. m. musculus/M. m. domesticus* F₁ hybrids.

2.4.3 Implications for Mammalian Speciation

In 1993, prior to the identification of CpG methylation as the molecular mechanism underlying genomic imprinting, Varmuza suggested that imprinted genes may have played a role in mammalian speciation by affecting hybridization between species. Disruption of genomic imprinting may lead to postzygotic isolation, traditionally held to be the result of one of three events: chromosomal arrangement differences, different ploidy levels, or allelic differences that are nonfunctional when brought together in hybrids (Coyne and Orr, 1998). Though loss of parental CpG methylation imprints does not alter the coding sequence at imprinted loci, the incompatibilities introduced between two parental genomes of different species or subspecies might still result in failure to properly establish imprinted gene expression in hybrid offspring. Such a severe impact at the developmental level could cause inviable hybrid embryos and eventual speciation.
Past analysis has indicated that early in mammalian evolution, the DMRs of imprinted loci diverged rapidly between species; following the establishment of specific lineages, such as mouse and rat, the evidence in support of rapid evolution is significantly lessened (Hutter et al., 2010). While further studies must be performed in humans and other mammalian lineages for a more comprehensive picture, these results underscore the extent to which imprinting mechanisms have become specific and fixed within taxa. The evidence obtained through this study suggests that even between *M. m. musculus* and *M. m. domesticus*, closely related subspecies whose sympatric ranges enable hybridization in their natural environments, F$_1$ hybrids experience disruption of CpG methylation patterns.

Should disruption of imprinting patterns become severe enough to impair the regulation of gene expression during embryogenesis of F$_1$ hybrids, early termination of pregnancy would be expected to occur. Early embryonic lethality resulting from hybridizations between closely related strains was reported prior to the identification of CpG methylation as the mechanism of imprinting: Renard and Babinet (1986) described termination of pregnancy at the blastocyst stage in crosses between *M. m. domesticus* DDK females and males of *M. m. domesticus* strains BALB/c or C57BL/6. Reciprocal crosses with DDK males and females of either BALB/c or C57BL/6 genotypes resulted in normal hybrid offspring. Only DDK females impregnated by a male of the DDK strain delivered normal litters, providing early evidence of a phenomenon now known as asymmetric postmating isolation (Turelli and Moyle, 2007). Sapienza *et al.* (1992) later mapped the gene responsible for the so-called “DDK syndrome,” *Ovum mutant*, to an
imprinted locus on mouse chromosome 11, determining that the syndrome arose from epigenetic incompatibilities during hybridization.

Since these early studies, researchers have continued to be surprised by the severe pathological consequences of lineage- and strain-specific imprinting mechanisms that are incapable of functioning against an unfamiliar genetic background. Vrana’s (2007) work with Peromyscus interspecific hybrids illustrates just how quickly epigenetic differences can be amassed between two closely related species. \( P. \) maniculatus and \( P. \) polionotus diverged in the Pleistocene and were initially thought to be insufficiently reproductively isolated to demonstrate significant epigenetic incompatibilities. Yet \( P. \) polionotus \( ♀ \times P. \) maniculatus \( ♂ \) \( F_1 \) hybrids show hypomethylation at normally silenced alleles and total loss of imprinting at multiple loci, resulting in fetal overgrowth and significant levels of embryonic lethality. As previously stated, \( M. \) m. \( m. \) musculus diverged from \( M. \) spretus only 1.5 million years ago, yet their hybrids, too, have demonstrated clear LOI (Shi et al., 2005). It can be concluded, then, that despite close evolutionary relationships between their parental strains, interspecific hybrids are not exempt from the accumulation of epigenetic disruptions (Zeh and Zeh, 2008). My results with \( M. \) intersubspecific hybrids further emphasize how finely tuned to a particular strain genomic imprinting mechanisms have become.
2.4.4 Genomic Imprinting and LOI in *Mus* Intersubspecific Hybrids

Whether these observed disruptions in CpG methylation at *H19*, *Rasgrf1*, and *Snrpn* result in pathological consequences in F\textsubscript{1} hybrids is unclear. In the laboratory, as in the natural hybrid zone, *M. m. musculus* and *M. m. domesticus* strains are capable of producing viable hybrid offspring; reproductive isolation is therefore considered to be incomplete (Guenet and Bonhomme, 2003). Their separation from a common ancestor is estimated to have occurred fairly recently. While I have established that the hybrid offspring differ from the parental strains at the molecular level, LOI cannot be accurately reported until aberrant biallelic expression is confirmed.

This is often accomplished through reverse transcriptase PCR (RT-PCR) and restriction fragment length polymorphism (RFLP) analysis (Oakey and Beechey, 2002). Following identification of SNPs or other differences between the coding regions of imprinted loci from two species or subspecies, RT-PCR is performed. Digestion of cDNA with restriction enzymes able to cleave only one of the two alleles allows RFLP gel analysis. As this general approach requires prior identification of sequence polymorphisms, its usefulness can be limited in studies employing mouse strains for which no SNPs are available (e.g., Tuskan *et al.*, 2008). While I identified no SNPs in the DMRs of these loci useful for this purpose, sequencing of the coding regions in *M. m. musculus* and *M. m. domesticus* individuals might provide sufficient sequence polymorphisms to perform such analyses and confirm LOI.
2.4.5 Hybrid Sterility and Haldane’s Rule

The possibility of hybrid sterility in *M. m. musculus/M. m. domesticus* hybrids resulting from LOI is of significant research interest. Despite the occurrence of F₁ hybrids in the natural hybridization zone between the two species, Britton-Davidian *et al.* (2005) reported sterile F₁ males as well as a significant degree of F₁ female infertility in laboratory crosses. However, their investigation into the genetic architecture of this hybrid sterility pointed to incompatibilities between two alleles which were not reported to demonstrate imprinted expression in the *M. m. musculus* and *M. m. domesticus* backgrounds.

To what extent epigenetic incompatibilities may have contributed to the observed incidence of hybrid sterility remains undetermined. Future studies may elucidate whether imprinted loci are widely involved in hybrid sterility in these and other *Mus* species and subspecies. Demonstrating interruption of epigenetic interactions may resolve, for example, why *M. spretus* experiences complete reproductive isolation from *M. m. domesticus* despite sympatric ranges in the Mediterranean, as previously described (Dejager *et al.*, 2009).

Also of interest is the possibility that mechanisms of hybrid sterility in laboratory strains may provide evidence for the cause of impaired reproduction in humans and for the increased risk of imprinting disorders in children born through ART. Mouse studies have already suggested that the inability to establish proper methylation at ICRs may underlie infertility and can be further compounded by aging of the oocytes (Lopes *et al.*, 2009). Mann *et al.* (2004) observed LOI at *H19* and *Igf2* following
preimplantation development in culturing with Whitten’s medium, commonly used during the manipulation of mammalian embryos for ART. Ovarian stimulation to obtain mouse oocytes for ART, as well as cryopreservation of oocytes and embryos, leads to LOI at multiple loci and an increased incidence of imprinting disruption (Market-Velker et al., 2010; Wang et al., 2010). Similar mechanisms may be responsible for the increased incidence of genetic abnormalities and epigenetic disorders in human pregnancies resulting from IVF (Liang et al., 2011).

Genomic imprinting may also help to explain, at least in part, deviations of mammalian taxa from Haldane’s rule. First described in 1922, Haldane’s rule posits that when hybridization between incompatible species results in deleterious effects, the heterogametic sex will be more frequently affected and is therefore more likely to be sterile, rare, or absent altogether. While this principle is borne out well by data collected from Drosophila and many other non-mammalian taxa, researchers have struggled for some time to understand why, in mammalian hybrids, male offspring often are not disproportionately affected (Jablonka and Lamb, 1991). It was eventually concluded that as purely genetic models were insufficient to explain these observations, it was necessary to take into account the developmental consequences of interspecies hybridization, including differences in epigenetic markers and the potential alterations in chromosomal structure they may cause. Recent studies do suggest that in Mus hybrid sterility is slightly more pronounced in male offspring, but the degree of female hybrid infertility is considerable (Britton-Davidian et al., 2005).
CHAPTER THREE: SELECTIVE PRESSURES AT IMPRINTED LOCI

3.1 Natural Selection and Genomic Imprinting

Rates of nucleotide change within lineages are controlled by the presence of natural selection acting either negatively or positively. Selection on protein-coding sequences results in differences between the rates of synonymous and nonsynonymous substitutions in the DNA sequence. Nonsynonymous changes alter the amino acid sequence; these mutations may deleteriously alter protein function and are thus more likely to be removed from the population by negative, or purifying, selection (Yang, 2006). Negative selection can act to constrain particular alterations in amino acid sequence, with its intensity varying depending on the number of nonsynonymous substitutions that can be tolerated by the affected protein without loss of function (Broughton and Reneau, 2006).

If, however, the change in amino acid sequence alters the protein encoded in such a way as to be advantageous to the organism, it will be favored by positive, or Darwinian, selection. Genes encoding such proteins will demonstrate an excess of nonsynonymous changes, indicating that positive selection favoring an advantageous new variation has fixed nonsynonymous substitutions at a higher level than synonymous substitutions (Anisimova et al., 2001). Across mammalian genomes, positive selection has been observed in genes that function in immunity, especially the major histocompatibility complex, and in olfactory genes (Emes et al., 2004). Upon finding very high rates of positive selection in human genes related to immune defense, Nielsen et al. (2005)
described positive selection as one result of the “coevolutionary arms race” waged by host cells and rapidly evolving pathogens such as viruses.

Analyses of selection on imprinted genes have presented a number of unexpected observations. As imprinted expression across mammals is strongly conserved, with the orthologs of most imprinted genes monoallelically expressed in other species, it was initially anticipated that the DMRs of imprinted loci would demonstrate similar constraints. Rather, the sequences of DMRs are often highly divergent between species and specific to particular lineages; one of the few conserved elements identified in the DMRs of multiple species is the presence of tandem repeats nearby or adjacent to the DMR (Hutter et al., 2010). Observation of tandem repeats is often combined with investigation for binding sites for CTCF and its cofactor Yin-Yang 1 (YY1), which have been shown to associate with the CpG-rich DMR sequences (Rodriguez-Jato et al., 2005; Yoon et al., 2005).

3.2 The $d_{S}/d_{S}$ Ratio in Molecular Evolution

Kimura (1983) asserted the importance of synonymous and nonsynonymous substitution rates in characterizing the dynamics of evolution at the molecular level in what has since become known as the neutral theory of molecular evolution. Selective pressures at particular loci can be estimated by examining the rate of substitutions as compared to expectations of neutral evolution. Researchers who investigate evolutionary change at the molecular level have reasoned that the pressures of selection at any particular locus will depend on how critical the protein encoded is to the proper physiological function of the organism (Wall et al., 2005).
In the absence of selective pressures, the rate of nonsynonymous substitutions ($d_N$) should not differ significantly from that of synonymous substitutions ($d_S$), which are assumed to have no effect: the amino acid encoded is not altered, so selection cannot intervene (Galtier and Duret, 2007; Higgs and Attwood, 2005). The $d_N/d_S$ ratio (also referred to as $\omega$) is therefore a property of a gene that reveals both how strongly selection is acting on that locus (the intensity) as well as the type of selection. If the ratio differs from 1.0, evidence exists in violation of neutral evolution (Smith and Hurst, 1998). A $d_N/d_S$ ratio greater than 1.0 suggests the presence of positive selective pressures acting at that particular locus. A very low ratio suggests purifying selection (McVean and Hurst, 1997; Tang and Wu, 2006). Rates of evolution will thus be higher in proteins that are less important to the organism as a result of accumulating synonymous changes (Wall et al., 2005).

While theory would dictate that $\omega$ must exceed 1.0 if positive selection is present, $\omega$ at loci demonstrating strong evidence of positive selection, including those genes functioning in immune defense, very rarely reaches 1.0. This is likely due to negative selection acting on portions of the genes which are not free to evolve rapidly without severely compromising protein function. In practice, then, $\omega$ values approaching 0.500 provide compelling evidence of positive selection, while values nearer 0.100 are taken to be indicative of negative selection.

Multiple methods have been developed to estimate the $d_N/d_S$ ratio. Some rely on phylogenetic trees to determine patterns of molecular change, while others use pairwise comparisons of aligned protein-coding sequences (Nei and Gojobi, 1986; Goldman and Yang, 1994). Pairwise estimates are considered to provide a stricter test of
positive selection across a gene, while phylogeny-based likelihood tests are favored to examine selective pressures at different codon positions within a gene for closely related species within the same genus (Jansa et al., 2003).

Hurst (1997) reviewed over a dozen hypotheses on the evolution of genomic imprinting, including defense against parthenogenesis, malignant ovarian trophoblastic disease, or chromosomal gain or loss; minimizing variance in rates of gene expression between offspring; and the conflict hypothesis. Though speculation continues, the conflict hypothesis remains the most favored model: it has been consistently favored by investigators in the field and is strongly supported by the overwhelming number of paternally-expressed imprinted genes which promote growth and maternally-expressed imprinted genes which suppress it. Impressively, this prediction was made before the functions of most imprinted genes were elucidated (Tilghman, 1999).

Though great progress has been made in cataloging the sequence features of imprinted loci, only a few analyses have been devoted to comparing molecular evolutionary rates of imprinted genes to those of non-imprinted genes (Hutter, 2010; McVean and Hurst, 1997; Smith and Hurst, 1998). Investigators initially expected that imprinted genes would be found to undergo rapid and antagonistic coevolution as a result of the relentless race between the two parental genomes to control resource utilization by the offspring. However, evidence from early studies of evolutionary rates in imprinted genes did not appear to support this model. McVean and Hurst (1997) predicted that rates of molecular evolution at imprinted loci would not differ significantly from the elevated $d_{\text{N}}/d_{\text{S}}$ ratios seen in immune genes. They were unable to demonstrate a statistically significant association, instead reporting that the $d_{\text{N}}/d_{\text{S}}$ values of the imprinted loci
surveyed suggested a rate of molecular evolution similar to those seen at neutral loci, including receptors such as the nicotinic acetylcholine receptor and insulin receptor.

Hurst and Smith (1998) argued that essential genes, defined as those for which knockouts are inviable, will demonstrate lower rates of molecular evolution. Grouping genes into essential, non-essential, and immune categories, they found that essential genes do indeed demonstrate $d_{NS}/d_S$ ratios significantly lower than those of non-essential genes, while immune genes demonstrated higher rates consistent with Darwinian selection driven by host-parasite coevolution. Moore and Mills (1999) concurred with Hurst and Smith (1998), stating that haploid expression from imprinted loci will expose these genes to severe effects from selection; maintenance of the hemizygous state would be expected to prevent the accumulation of mutations leading to the expression of a deleterious allele. Nonsynonymous mutation rates at imprinted loci would therefore be expected to remain low (Moore, 2001).

3.3 Bioinformatic Analysis of Molecular Evolution

Following the remarks of Moore and Mills (1999), Zeh and Zeh (2000) also posited that alterations in the amino acid sequence of the expressed allele would be expected to be removed rapidly by selection at imprinted loci. Furthermore, if imprinted genes are involved in mammalian speciation, disruption of sequences, like loss of appropriate epigenetic modifications, would be expected to result in barriers to hybridization. Analysis of molecular evolutionary rates at these and other protein-coding sequences for imprinted loci in *Mus* would be expected to demonstrate evidence of purifying selection.
This expectation is in contrast to the earlier predictions of strong Darwinian selection expected under a model of intragenomic antagonistic coevolution between female and male genomes. Rather, I suggest that statistical analysis of molecular evolutionary rates will present evidence of purifying selection. Specifically, $d_{s}/d_{\bar{s}}$ values derived from protein-coding sequences of imprinted genes are expected to differ significantly both from loci undergoing neutral selection and those undergoing strong positive selection as a result of antagonistic coevolution. Values are not expected to differ significantly from the low rates of evolution seen in essential genes. Comparison of $\omega$ for mouse-human orthologous sequences with those of the mouse and human genomes overall, performed by Hutter et al. (2010), provided the median value of $\omega$ (0.120) for imprinted loci. Their results proposed that while imprinted sequences are relatively free to accumulate synonymous mutations, with $d_{s}$ exceeding 0.600, the majority of $\omega$ values remain below 0.25, indicating purifying selection (Rat Genome Sequencing Project Consortium, 2004).

The three protein-coding loci chosen for study by bisulfite sequencing ($Igf2r$, $Rasgrf1$, and $Snrpn$) were investigated. The sample size was expanded by including additional loci known to be imprinted in both $M.\text{musculus}$ and $Rattus\text{norvegicus}$: ankyrin repeat and SOCS box protein 4 ($Asb4$), carboxypeptidase A4 ($Cpa4$), type III iodothyronine deiodinase ($Dio3$), Distal-less homeobox 5 ($Dlx5$), growth factor receptor-bound protein 10 ($Grb10$), insulin-like growth factor 2 ($Igf2$), insulin 1 ($Ins1$), insulin 2 ($Ins2$), and ubiquitin-specific peptidase 29 ($Usp29$) (Hutter, 2010; McVean and Hurst, 1997). The final set included 12 imprinted loci.
3.4 Methods and Materials

Rates of molecular evolution at imprinted loci were compared against three control sets in order to evaluate selective pressures. The first control set is a selection of 12 non-essential, non-immune genes undergoing neutral evolution (Hurst and Smith, 1999). Loci include: ciliary neurotrophic factor (Ctnf), dopamine receptor D2 (D2dr), dopamine receptor D4 (Drd4), glial fibrillary acidic protein (Gfap), metabotropic glutamate receptors 1 and 2 (Grm1, Grm2), gelatinase A (Mmp2), opioid receptor μ-1 (Oprm1), preproenkephalin (Penk), peripheral myelin protein 22 (Pmp22), synapsin 1 (Syn1), and tissue inhibitor of metalloproteinase 1 (Timp1).

The second set consists of 12 immune genes which undergo rapid Darwinian selection in Mus and Rattus as a result of coevolution with pathogens (McVean and Hurst, 1997). Loci include serum amyloid P component (Apcs); antigens CD2, CD8α, CD8β, CD48, and CD74; C-reactive protein (Crp); interleukins 2, 3, 4, and 9 (IL2, IL3, IL4, IL9); and thymus antigen 1 (Thyl).

The third set consists of 12 essential loci under strong purifying selection in Mus and Rattus (Hurst and Smith, 1999). Loci include anti-Müllerian hormone (Amh), Bcl2-associated X protein (Bax), brain-derived neurotrophic factor (Bdnf), early growth response 1 (Egr1), estrogen receptor 1 (Esr1), homeobox A1 (Hoxa1), leukemia inhibitory factor (Lif), mannose acetylgalucosaminyltransferase 1 (Mgat1), progesterone receptor (Pgr), prolactin receptor (Prlr), survival motor neuron (Smn), and vascular endothelial growth factor A (Vegfa).
Protein-coding sequences for all three sets of mouse genes were obtained from the RefSeq database of the National Center for Biotechnology Information. *R. norvegicus* sequences orthologous to *M. musculus* genes were identified through HomoloGene. Orthologous sequences are used to control for the date of speciation events: orthologs resulting from the same speciation event have the same divergence date, and differences in \( d_N \) and \( d_S \) can therefore be taken to reflect the number of substitutions per site, per year (Douret and Mouchiroud, 2000). See Appendix B for GenBank accession numbers of mouse and rat orthologous sequences (Tables 5-8).

Alignments of rat and mouse cDNA sequences were performed using ClustalW. Molecular Evolutionary Genetics Analysis (MEGA) 5.0 determined rates of synonymous and nonsynonymous substitution by modified Nei-Gojobori method with Jukes-Cantor correction (Nei and Gojobori, 1986; Tamura et al., 2011). Mann-Whitney \( U \)-tests and Kruskal-Wallis analysis were performed with PASW Statistics 18 (IBM, Somers, NY). Sliding window analysis was performed with DNAsp 5.0 (Librado and Rozas, 2009) with 200-bp increments.

3.5 Results

The Mann-Whitney \( U \)-test provides non-parametric analysis of two independent groups of values. Under a null hypothesis (\( H_0 \)) of neutral evolution, both sets of values could have come from the same population; the separation observed between the values could be the result of chance alone. If there are no significant differences between the experimental group, the ratios of nonsynonymous to synonymous substitution in the 12
imprinted loci (Table 9, Appendix B), and the control group of non-essential loci (Table 10, Appendix B), $H_0$ cannot be rejected. If $U$-test results provide a $p$-value indicating that the differences observed in the values are unlikely to be the result of chance alone, evidence will exist in support of the alternative hypothesis ($H_A$) of selective pressures at imprinted loci. The test was then repeated to determine whether imprinted loci differ significantly from immune loci and essential loci.

Following analysis of ClustalW alignments with MEGA 5.0, values for $d_N$ and $d_S$ were obtained for each locus (see Tables 9-12, Appendix B). These values were plotted to visualize general trends between imprinted loci and the three control groups (Figure 9).

![Figure 9](image.png)

**Figure 9: Number of synonymous ($d_N$) and nonsynonymous ($d_S$) substitutions per 100 nucleotides.** Initial linear regression indicated that the slope of imprinted loci (0.0912, diamonds) was closest to that of essential loci (0.1246, triangles), while non-essential loci (0.6365, squares) and immune loci (0.4857, crosses) were steeper.
Imprinted, non-essential, and essential $d_N$ and $d_S$ rates were largely clustered together. Correcting for the outlier ($Syn1$) in the values for non-essential loci resulted in a much lower slope (0.0649) closer to that of imprinted loci, suggesting that with the exception of one locus, $\omega$ values do not differ significantly between the two groups. Rates of nonsynonymous and synonymous substitution were then used to determine $\omega = d_N/d_S$ at imprinted loci, non-essential loci, essential loci, and immune loci.

Mean $\omega$ for imprinted genes ($0.135 \pm 0.034$) is below 0.25, as are all but two individual $\omega$ values, consistent with the findings of Hutter et al. (2010) and suggestive of the presence of purifying selection in this group. Mean $\omega$ for imprinted loci was smaller than the mean ratio at non-essential loci ($\omega = 0.151 \pm 0.053$) and immune loci ($\omega = 0.676 \pm 0.074$) and larger than that of essential loci ($\omega = 0.130 \pm 0.020$).

Figure 10: Mean ratios of nonsynonymous and synonymous substitution rates. MEGA 5.0 was used to determine $d_N$ and $d_S$ rates. Error bars represent ± standard error of the mean. All four sets included 12 loci.
However, analysis of $\omega$ values with Mann-Whitney $U$-tests found that imprinted loci do not differ significantly from non-essential loci or essential loci. They were found to differ significantly only from immune loci.

Table 3: Mann-Whitney $U$-test statistics for comparison of $d_S/d_S$ ratios. Values for imprinted loci were grouped ($n = 12$) and compared with the following control sets: essential loci ($n = 12$), immune loci ($n = 12$), and non-essential loci ($n = 12$). Tests were performed using PASW Statistics 18.

<table>
<thead>
<tr>
<th>Group</th>
<th>U</th>
<th>Z</th>
<th>$p$ (exact)</th>
<th>$p$ (asymp.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Essential</td>
<td>59</td>
<td>-0.772</td>
<td>0.478</td>
<td>0.470</td>
</tr>
<tr>
<td>Immune</td>
<td>0.00</td>
<td>-4.157</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Non-Essential</td>
<td>68</td>
<td>-0.231</td>
<td>0.843</td>
<td>0.817</td>
</tr>
</tbody>
</table>

These results indicate that $d_S/d_S$ at imprinted loci differs significantly from $d_S/d_S$ of immune loci and therefore do not support a model of positive selection. Without statistical evidence indicating that imprinted loci differ significantly in their molecular evolutionary rates from non-essential loci, however, it is not possible to support the model of negative selection.

A Kruskal-Wallis $H$-test was then performed to investigate for significant differences between imprinted, essential, and non-essential $\omega$ values. Like the Mann-Whitney $U$-test, the Kruskal-Wallis $H$-test operates on the null hypothesis that the ranks of the values within each group do not differ. I failed to reject this null hypothesis ($H = 0.96, p = 0.619$); it is therefore not possible to draw conclusions about whether
molecular evolutionary rates at imprinted loci are subject to negative selection or are neutrally evolving based on these data.

Sliding window analysis was performed to determine whether the overall $d_N$ value is biased due to regions of high nonsynonymous substitution at the locus for which the highest $\omega$ value was obtained, \textit{Rasgrf1}. Increments of 200 bp were used as calculations of $d_N/d_S$; smaller increments (100 bp or fewer) have been shown to be error-prone and result more frequently in false-positive $d_N$ peaks, while larger increments tend to underestimate the occurrence of $d_N$ peaks (Parmley and Hurst, 2007). A $d_N/d_S$ peak, as defined by Parmley and Hurst (2007) as a maximum point with at least six flanking windows having lower ratios, was identified between 2400-2600 bp.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{rasgrf1 LOCUS IN EIGHT SPECIES}
\caption{	extbf{Figure 11: Sliding window analysis of \textit{Rasgrf1} loci in four species.} Analysis performed with DNAsp v. 5.0. Windows occur in 200-bp segments. The conserved peak occurring at approximately 2600 bp (approximately 866 amino acids) is defined as significant under the criteria of Parmley and Hurst (1997), which require at least six windows on either side of the peak demonstrating a lower ratio.}
\end{figure}
In addition to the mouse-rat alignment, mouse-human (NP_002882), mouse-chimpanzee (XP_001153586), and mouse-dog (XP_545892) orthologous Rasgrf1 coding sequences were aligned by ClustalW and analyzed in 200-bp windows. Mouse-rat, mouse-chimpanzee, and mouse-human sequences obtained through HomoloGene include the full coding sequence with the stop codon present at the final three bases. All three demonstrate maxima at approximately 2600 bp. The maximum $\omega$ for the mouse-dog comparison occurs at approximately 2200 bp (733 amino acids) with a second, slightly smaller peak at approximately 3200 bp. The dog sequence obtained through HomoloGene does not have a stop codon present in the final three base pairs and does not appear to represent the full coding sequence. Thus, it does not provide a complete characterization of $d_N/d_S$ ratios across the full canine Rasgrf1 coding sequence.

3.6 Discussion

These data are in agreement with the findings of Hutter (2010) and O’Connell et al. (2010) overall, with Mann-Whitney $U$-tests indicating that the imprinted and immune $d_S/d_S$ values differ significantly. However, as previously stated, several imprinted loci demonstrate surprisingly elevated rates of nonsynonymous substitution. $Ins1$, **Rasgrf1**, and $Usp29$ have $d_N$ values of $0.034 \pm 0.014$, $0.039 \pm 0.010$, and $0.050 \pm 0.005$, respectively. $Ins1$ and $Usp29$ do not demonstrate $d_S/d_S$ exceeding 0.25, however, and are therefore still consistent with the broad definition of purifying selection.
3.6.1 Site-Specific Positive Selection at *Rasgrf1*

Two imprinted loci in this data set, *Igf2r* and *Rasgrf1*, demonstrated $\omega$ exceeding 0.25. McVean and Hurst (1997) used sliding window analysis in 240-base-pair blocks to further examine $\omega$ across *Igf2* and *Igf2r* in order to determine whether elevated rates of nonsynonymous substitution might result from positive selection acting on domains required for binding. This is termed site-specific positive selection and was recently reported in the imprinted loci *Osbpl5* and *Gnaslx*, in which regions under significant Darwinian selection corresponded to evolutionarily conserved domains required for protein-protein interactions (O’Connell *et al.*, 2010). McVean and Hurst (1997) reported no evidence for site-specific selection at binding sites in *Igf2* or *Igf2r*, however, stating that characterization of $d_N/d_S$ ratios in the binding region did not support positive selection. Rather, the $d_N/d_S$ value was found to be lower in this region, suggesting the action of purifying rather than positive selection; no support was reported for the action of positive selection across the entire locus.

Smith and Hurst (1999) later suggested that the increased incidence of CpG dinucleotides at the DMR within the coding region of *Igf2r* and the propensity of methylated CpGs toward CpG ↔ TpG substitutions might account somewhat for the variations in $d_S$ across the locus. Further analysis at the *Igf2r* locus did show that positive selection is indicated in the region of the *Igf2r* signal sequence, suggesting that *Igf2r* may have coevolved with another protein product—perhaps encoded by a paternally-expressed locus—which controls its cellular localization.
I performed sliding window analysis to further explore nonsynonymous substitution rates across the other locus which demonstrated $\omega$ in excess of 0.25, Rasgrf1. As the DMR is located some 30 kb upstream of the coding region, the presence of CpG dinucleotides would not be expected to contribute to variations in $d_S$ leading to the peaks and troughs visible in Figure 11. Rather, the conserved peak in $\omega$ across all four sequence comparisons is considered significant under Parmley and Hurst’s (1997) criteria and is suggestive of site-specific positive selection at the Rasgrf1 locus. The peak is strongest for the mouse-rat lineage, possibly due to the greater number of synonymous changes that have accumulated between the mouse and its more distant relatives such as chimpanzee and human, lowering $\omega$ across the coding sequence.

O’Connell et al. (2010) found evidence of positive selection acting in 14 imprinted loci, including Rasgrf1 and Igf2, but stated that selective pressures appear to vary across lineages. Igf2, for example, only demonstrated regions of positive selection in the mouse and opossum. To make matters more complex, they also reported 33 cases of known imprinted loci where no orthologs demonstrate positive selection. They summed up their findings as a lack of correlation between known imprinting status and evidence of positive selection in mammals.

A more complete set of orthologous protein-coding data might be able to further refine patterns of $d_S/d_S$ throughout these loci. In particular, it would be useful to have sequences for more than four confirmed Rasgrf1 mammalian orthologs. Sequence quality and completeness may also affect the results obtained for Figure 11, as it is ideal to use an annotated sequence for which start and stop codons are known. The sequence obtained
for dog (*Canis familiaris*), for example, does not contain a stop codon at the 3’end and is therefore an incomplete coding sequence.

### 3.6.2 Functions of Rasgrf1 in Mammalian Physiology

The RAS proteins are the founding members of the superfamily of small GTP-binding proteins referred to as RAS-like GTPases. All known family members function in signal transduction. Specifically, they assemble complexes at the cell membrane responsible for activating further signaling and ensuring cell cycle progression (Rajalingam *et al*., 2007). Investigations of RAS signaling in the neurons of adult mice have shown that individuals deficient in a number of proteins in the RAS family are unable to properly employ mitogen-activated protein kinase/extracellular signal-regulated kinase (MAPK/ERK) signaling (Tian *et al*., 2004). Among other effects, they were found to have impaired learning abilities and memory, linked to both the amygdala and the hippocampus (Brambilla *et al*., 1997; Giese *et al*., 2001). Proteins known to interact with RAS family members are characterized by RAS binding domains (RBD), of which three distinct forms have been identified (Hermann, 2003).

Within the last decade, Rasgrf1 has been shown to be required for the initiation of signaling cascades at NMDA receptors (NMDARs) (Krapivinsky *et al*., 2003). Activation of the MAPK/ERK signaling pathway is NMDA-dependent and requires an increase in intracellular Ca\(^{2+}\). The effects of the Ca\(^{2+}\) increase have been shown to be strictly spatially regulated, affecting only those synaptic NMDARs in the immediate vicinity (Hardingham *et al*., 2001). Increased Ca\(^{2+}\) has been shown to act as an “on” switch for
the ERK1/2 pathway, which is closely coupled to NMDARs and eventually targets the transcription factor CREB (Tian et al., 2004).

Rasgrf1 interacts directly with NR2B, a subunit of certain NMDARs, as demonstrated by co-immunoprecipitation of the Rasgrf1-NR2B complex with NR2B antibody (Krapivinsky et al., 2003). Investigating the relationship further, the authors found that the region necessary and sufficient to bind NR2B is located between amino acids 714-913. Notably, the peak $d_S/d_S$ value I obtained for sliding window analysis occurs at approximately 800-866 amino acids and is contained within the range required for binding. The discovery of this interaction provided a model for MAPK/ERK activation: Rasgrf1 provides a kind of sensor for Ca\(^{2+}\) concentration in the vicinity of NMDARs. It becomes physically confined to the NMDA channel by its association with NR2B. Only NR2B-containing NMDARs will be activated by the Rasgrf1-NR2B complex, the formation of which is dependent on the presence of increased Ca\(^{2+}\).

This pathway is critical for neuronal signaling and plasticity (Kim et al., 2005; Massey et al., 2004; Zhao et al., 2005). Rasgrf1 is among the very few imprinted genes whose expression pattern alters throughout development: though monoallelic expression is retained in the majority of tissues, by weaning its expression in the brain becomes biallelic (Drake et al., 2009). This requirement for strictly regulated patterns of expression suggests discrete, highly significant effects on neurological and behavioral development, especially given that the switch to biallelic expression is limited to mature neuronal cells and is absent from the glia (Davies et al., 2005; Zippel et al. 1997). Biallelic expression in the brain and the maintenance of elevated levels of protein near
synaptic sites—confirmed in the work of Sturani et al. (1997)—are consistent with the critical role of the Rasgrf1-NR2B complex in signal transduction.

Thus it seems possible that positive selection and coevolution between these two proteins may have resulted in the switch from imprinted to biallelic Rasgrf1 expression in the mouse brain. This raises the question of whether the association between these two proteins can be confirmed to be evolutionarily recent and has reestablished biallelic Rasgrf1 expression, or whether their association is of long standing and has allowed Rasgrf1 to resist the pressure toward imprinting in a tissue-specific manner. Davies et al. (2004) discussed the “leaky” expression of Rasgrf1 seen at the silenced maternal allele in the newborn mouse brain; it remains unclear whether or not silencing was complete at some point in the mammalian lineage.

These inferences have been drawn tentatively because a larger data set containing additional mammalian Rasgrf1 orthologs is absolutely critical to obtaining a more comprehensive and accurate picture. Before a definitive statement of site-specific positive selection can be made, confirmation of the expression patterns of Rasgrf1 and coding sequences from a large number of mammalian species must be obtained. Future analyses might also consider the selective pressures at the locus encoding NR2B as well as any site-specific variations in $d_N/d_S$. To confirm positive selection and coevolution, $d_N/d_S$ peaks corresponding to the Rasgrf1-NR2B binding site would have to be identified in both proteins for closely related taxa.
3.6.3. The Absence of Overall Positive Selection

With statistical analysis suggesting that molecular evolutionary rates at imprinted loci differ significantly from those of immune loci, it is reasonable to conclude that imprinted loci as a whole are likely not subjected to positive selection. I found no evidence in favor of such rapid molecular evolution in the data set. On this point, my conclusions concur with O’Connell et al. (2010), who stated that widely applicable support for antagonistic coevolution in imprinted protein-coding loci resulting from the putative parental conflict does not currently exist.

These results are not incompatible with the conflict hypothesis of the evolution of genomic imprinting. The process of imprinting may be the result of selection acting on the competition between the maternal and paternal genomes while the gene products—the vast majority of which are critical for placental and fetal development—are not altered. Dosage control rather than evolution of the nucleotide sequence can accomplish this aim (Drake et al., 2011). This would offer one possible explanation of why we do not see the signature of widespread positive selection. Occurrences such as the \( d_{\text{s}}/d_{\text{S}} \) peak of Rasgrf1 may reflect associations with other proteins that predate the origins of imprinting in placental species, as Rasgrf1 orthologs in non-mammals have also shown evidence of positive selection (O’Connell et al., 2010).

Before these questions can be answered with accuracy, however, the genomic imprinting status of a wide array of lineages must be determined. It also remains to be determined whether duplication of an imprinted locus might allow one paralog to undergo rapid evolution while the other remains under purifying selection. Described by
the model of neo-functionalization, the presence of paralogous genes resulting from a duplication event would, in principle, allow one copy to evolve toward a new function or role in a different cellular pathway while experiencing a period of relaxation in selection, similar to that experienced by pseudogenes; it might retain some or none of its original function (He and Zhang, 2005). This was demonstrated by Spillane et al. (2007) in the adaptation of the imprinted plant gene Medea in the Arabidopsis lineage, whose $\omega$ value is 1.68. While convincing evidence does not yet exist, further studies may show that imprinted genes demonstrating evidence of positive selection may are the result of gene duplication events (Hutter et al., 2010). Once identified within the genome, their paralogs would be expected to demonstrate strict purifying selection, as they were required to retain their original physiological functions.

It is important to note that while positive selection may not be acting on imprinted loci as a whole, these data do not distinguish whether neutral or negative selective pressures are at work. As originally proposed, $\omega$ is used to distinguish those amino acid changes that result from selective pressures from those changes that are simply the background nucleotide mutation rate, $\pi$ (Xing and Lee, 2006). Neutral selection is assumed to occur at unity ($d_N = d_S$), with deviations in either direction indicative of selective pressures, but $\omega$ values equal to precisely 1.0 in neutrally evolving genes have not been convincingly demonstrated.

As suggested by the wealth of studies demonstrating the presence of splice sites and motifs required for secondary RNA structure within coding regions, the reality of selection at synonymous sites may be far more complex (Parmley and Hurst, 2007). A number of critics have already proposed that features other than amino acid sequence are
subject to selective pressures and thus contribute to molecular evolution of genes (Chamary et al., 2006; Chamary and Hurst, 2005; Duan et al., 2003; Sloan and Taylor, 2010). Mounting evidence for selection against altered sequences at RNA splicing motifs, including substantially lowered $d_S$ rates, has led to the “RNA selection pressure” model (Baek and Green, 2005; Xing and Lee, 2006).

Other authors have argued that translational selection arises from codon usage bias, the result of differing translational efficiencies of codons during peptide synthesis. Though already accepted to occur in yeast genomes, this principle is still being explored in mammalian genomes (Akashi, 2003; Lavner and Kotlar, 2004). Adjustments are now made to account for restrictions on evolution at silent sites due to codon bias in studies of gene divergence in yeast, based on the observations of four species in the genus Saccharomyces (Hirsh, 2005).

As mammalian genomic studies progress, similar models for molecular evolution in more complex genomes may be developed. This will require a wealth of data in order to provide a more comprehensive picture of selective pressures, especially potential selection at silent sites. In the meantime, any conclusions drawn about $d_N/d_S$ ratios, on both the site-specific level and across entire loci, will continue to be limited by the lack of knowledge about whether mutation rates vary within particular regions of mammalian genomes (Castane et al., 1997). These steps are critical for researchers to distinguish whether $d_S$ values may be accurately treated as the proxy of neutral substitution rates and to avoid a skewed picture of the effects of natural selection. It is hoped that future algorithms will be capable of identifying non-coding DNA critical for transcriptional
regulation and sequences subject to translational selection from those regions in which synonymous changes truly reflect neutral evolution.
CHAPTER FOUR: SUMMARY AND CONCLUSIONS

Characterizing the relationship between molecular regulation and evolution of imprinted loci has raised a number of questions regarding both the evolution of mammalian lineages and susceptibility to aging and disease. As more data become available from genome sequencing, this field presents researchers with numerous challenging questions and contradictions. While multiple lineages demonstrate severe pathologies associated with LOI, it does not appear that CpG methylation dynamics are highly conserved across species (Beaujean et al., 2004). Given the variation in the locations of DMRs and ICRs and lineage-specific methylation dynamics, genomic imprinting has likely played an influential role in the evolutionary history of those species it affects.

In particular, the results obtained through bisulfite sequencing of *M. musculus* imprinted loci emphasize the importance of considering epigenetic markers as a source of postzygotic incompatibilities, even in two such closely related species as *M. m. musculus* and *M. m. domesticus*. While hybridizing mammalian species may not differ in terms of ploidy level or overall chromosome structure, genomic imprinting introduces an additional level of complexity: strain-specific differences in methylation at the DMRs of imprinted loci may affect placental and fetal growth following interspecific mating, even in instances of closely related species whose alleles at these loci do not differ significantly in their coding sequences.

Studies of genomic imprinting have led researchers to reconsider the process of evolution in the mammalian lineage. Rather than requiring drastic alterations in gene products—i.e., changes in coding sequences of critical genes—the diversity of
mammalian species may instead be the result of reprogramming, in part via genomic imprints, the spatial and temporal gene expression of offspring. Using the analogy of computer hardware and software, one of the leading researchers in the field, Dr. Randy Jirtle, likened the occurrence of evolution in the mammalian lineage to “rewriting the software of the computer” (Hunter, 2007). Because CpG methylation provides a reversible epigenetic mark, it is itself reprogrammable and would enable gene expression to be modified in such a way as to be specific in time and to particular tissues.

Like changes in computer software, alterations in genomic imprinting can come with a price. My results further highlight the need for investigation into the pathological effects of LOI resulting from mating between closely related species. When considering imprinting disruptions in humans, it is also important to further investigate the potential for raising the risk of imprinting disorders through ART. Niemitz and Feinberg (2004) issued a “call for investigation” after the results of a series of studies—including a comprehensive analysis by Schieve et al. (2002) of 42,000 children of couples who employed ART in the United States—indicated that the incidence of imprinting disorders is significantly increased in ART-conceived children (Cox et al., 2002; DeBaun et al., 2003; Hansen et al., 2002; Maher et al., 2003). Though epigenetic defects have been appearing for years in cattle and sheep conceived through ART in the animal husbandry industry (McEvoy et al., 2000), exploration into the possibility of a causal relationship between ART and imprinting defects in humans has begun only recently.

Mouse models of LOI will be critical to answering many remaining questions about genomic imprinting and its effects on mammalian reproduction and speciation. Given the evidence favoring a role for genomic imprinting in mammalian speciation, how
influential have imprinted genes been in determining reproductive isolation between species? Epigenetic alterations, including disruption of genomic imprinting patterns, may impact reproductive fitness independently of ART. Studies have already established a striking link between LOI at the H19 and Snrpn DMRs and moderate to severe oligozoospermia and abnormal chromatin packaging in human sperm (Hammoud et al., 2010; Marques et al., 2004; 2008). Human females demonstrate a natural but dramatic decline in reproductive fitness past the mid-30s. Once ascribed to the depletion of oocytes, this decline may also be related to degradation of epigenetic markers and compromised oocyte quality (Hamatani et al., 2004; Lopes et al., 2009).

To what extent, then, do epigenetic incompatibilities between maternal and paternal genomes underlie infertility, and does manipulation of the parental germ lines in vitro increase the susceptibility of any future offspring to imprinting disorders? It has already been argued that in human couples struggling to conceive, LOI should be viewed not as a side-effect but as a candidate for the cause of infertility itself (Marques et al., 2008). Past studies in mice and rats have demonstrated that treatment with the hypomethylating anticancer compound 5-azacytidine, which incorporates into replicating DNA and binds DNMTs as covalent adducts, leads to globally decreased DNMT activity, genomic hypomethylation, and impaired spermatogenesis (Kelly et al., 2003). Control treatments with the nonhypomethylating analog 6-azacytidine have no such effects, suggesting that it may be reduced expression or activity of DNMTs that leads to infertility in certain individuals. Furthermore, researchers must consider whether the introduction of epigenetic instability through ART will lead to more
serious complications for offspring into adulthood, including an increased risk of developing cancer.

Studies continue to reveal new and fascinating implications for the role of genomic imprinting in evolution. As the disciplines of genomics and epigenomics mature, it will be possible to reach a greater understanding of the origin and purpose of genomic imprinting, its role in our evolutionary history and physiology, and our relationships to other mammalian taxa. Epigenetic control paints a picture of genomic regulation that is far from static. Rather, it is becoming increasingly apparent that genomes are dynamic and exceedingly more complex than early researchers seeking to synthesize evolutionary biology and genetics could possibly have predicted.
REFERENCES


Smith, N.G.C., Hurst, L.D. (1999). The causes of synonymous rate variation in the rodent genome: Can substitution rates be used to estimate the sex bias in mutation rate? Genetics. 152, 661-673.


APPENDIX A

DMR sequences were obtained for \textit{H19} (U19619), \textit{Igf2r} (L06446), \textit{Rasgrf1} (NT_039476), and \textit{Snrpn} (AF081460) from GenBank. Primer sequences specific to the mutagenized DNA were described previously (Davis \textit{et al.}, 2000; Dockery \textit{et al.}, 2009; Hiura \textit{et al.}, 2006; Lucifero \textit{et al.}, 2002). Primers for negative controls were designed with Primer3 software (Rozen and Skaletsky, 2000) to amplify the DMRs from unconverted gDNA sequence.

\textbf{Table 4: Primer sequences for amplifying negative controls.} Sequences are specific to the original (unconverted) gDNA sequence and were used to assay the same regions as primers specific to the post-conversion (mutagenized) gDNA.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Primer</th>
<th>Orientation</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{H19}</td>
<td>H19D-FNC</td>
<td>Forward</td>
<td>5′-AAAATCGATTGCGCCAAAA-3′</td>
</tr>
<tr>
<td>\textit{H19}</td>
<td>H19D-RNC</td>
<td>Reverse</td>
<td>5′-CATTCCATGATCACCACCA-3′</td>
</tr>
<tr>
<td>\textit{Igf2r}</td>
<td>Igf2rL-FNC</td>
<td>Forward</td>
<td>5′-GCCCTCTGAATCCTCCTGTC-3′</td>
</tr>
<tr>
<td>\textit{Igf2r}</td>
<td>Igf2rL-RNC</td>
<td>Reverse</td>
<td>5′-GGATTCGAAGGGTTCTGTGA-3′</td>
</tr>
<tr>
<td>\textit{Rasgrf1}</td>
<td>Rasgrf1D-FNC</td>
<td>Forward</td>
<td>5′-CATATGACTCCGCCTGCTG-3′</td>
</tr>
<tr>
<td>\textit{Rasgrf1}</td>
<td>Rasgrf1D-RNC</td>
<td>Reverse</td>
<td>5′-GTGTGACATTTTGGGTTGTC-3′</td>
</tr>
<tr>
<td>\textit{Snrpn}</td>
<td>SnrpnH-FNC</td>
<td>Forward</td>
<td>5′-TGTGATGCTCCTCGCAATCTTTT-3′</td>
</tr>
<tr>
<td>\textit{Snrpn}</td>
<td>SnrpnH-RNC</td>
<td>Reverse</td>
<td>5′-GGATTCGACCACAAATTCTCAA-3′</td>
</tr>
</tbody>
</table>
Each PCR reaction to amplify converted gDNA was performed with ZymoTaq PreMix in 50.0-μL volumes with 5.0 μL bisulfite-treated DNA, 2.5 μL of each primer, 0.5 μL dNTPs, 25.0 μL buffer, 1.5 units Taq polymerase, and sterile distilled water. PCR reactions for negative controls were performed with Invitrogen Platinum Taq in 25.0-μL volumes with 1.0 μL gDNA, 1.0 μL of each primer, 0.5 μL dNTPs, 2.5 μL buffer, 0.5 μL MgCl₂ (25 mM), 1.0 unit of Taq polymerase, and sterile distilled water.

PCR with ZymoTaq PreMix was performed under the following conditions: 15 min at 95°C; 35 cycles of 30 s at 94°C, 1 min at 55°C, and 1 min 30 s at 72°C; and 7 min final extension at 72°C. Amplification of negative controls was performed with 5 min initial denature at 94°C; 30 cycles of 30 s at 94°C, 30 s at 60°C, and 30 s at 72°C; and 7 min final extension at 72°C. Products were isolated by agarose gel electrophoresis and all bands were extracted using the Qiagen QIAquick Gel Extraction Kit (Qiagen, Valencia, CA).

Sequencing was performed with 5.0 μL template DNA (bisulfite-treated or untreated for negative controls), 2.0 μL forward or reverse primer (at 1 mM), and BigDye Terminator v3.1 (2.0 μL buffer and 2.0 μL BigDye Terminator per sample) with provided pGEM control DNA, which was amplified with M13 primers.

The effectiveness of the Zymo EZ DNA Methylation Kit was determined with control mouse gDNA obtained through Zymo Research. Amplification with provided primers confirmed the successful detection of methylated CpGs within the control sequence, while unmethylated cytosines are subject to bisulfite conversion (see Figure 12).
Figure 12: Zymo Research methylated mouse gDNA control. PCR amplification was performed with primers provided by Zymo Research. Direct nucleotide sequencing confirmed that methylated CpGs were successfully detected while all other cytosine nucleotides were converted and detected as thymine.
### APPENDIX B

#### Table 5: Mouse and rat orthologs of imprinted loci.

<table>
<thead>
<tr>
<th>Imprinted Locus</th>
<th>GenBank Accession No. (<em>Mus</em>)</th>
<th>GenBank Accession No. (<em>Rattus</em>)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Calcr</em></td>
<td>NP_031614.2</td>
<td>NP_446268.2</td>
</tr>
<tr>
<td><em>Cdkn1c</em></td>
<td>NP_034006.3</td>
<td>NP_001028930.1</td>
</tr>
<tr>
<td><em>Cpa4</em></td>
<td>NP_082202</td>
<td>NP_001102816.1</td>
</tr>
<tr>
<td><em>Dcn</em></td>
<td>NP_031859.1</td>
<td>NP_077043.1</td>
</tr>
<tr>
<td><em>Grb10</em></td>
<td>NP_034475.2</td>
<td>NP_001102563.1</td>
</tr>
<tr>
<td><em>Igf2</em></td>
<td>NP_034644.2</td>
<td>NP_113699.2</td>
</tr>
<tr>
<td><em>Igf2r</em></td>
<td>NP_034645.2</td>
<td>NP_036888.1</td>
</tr>
<tr>
<td><em>Ins1</em></td>
<td>NP_032412.3</td>
<td>NP_062002.1</td>
</tr>
<tr>
<td><em>Ins2</em></td>
<td>NP_032413.1</td>
<td>NP_062003.1</td>
</tr>
<tr>
<td><em>Rasgrf1</em></td>
<td>NP_001034744.1</td>
<td>NP_001099223.1</td>
</tr>
<tr>
<td><em>Snrpn</em></td>
<td>NP_001076430.1</td>
<td>NP_112379.1</td>
</tr>
<tr>
<td><em>Usp29</em></td>
<td>NP_067298.2</td>
<td>NP_001101935.1</td>
</tr>
</tbody>
</table>

#### Table 6: Mouse and rat orthologs of non-essential, non-immune loci.

<table>
<thead>
<tr>
<th>Non-Essential Locus</th>
<th>GenBank Accession No. (<em>Mus</em>)</th>
<th>GenBank Accession No. (<em>Rattus</em>)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Ctnf</em></td>
<td>NP_740756.1</td>
<td>NP_037298.1</td>
</tr>
<tr>
<td><em>D2dr</em></td>
<td>NP_034207.2</td>
<td>NP_036679.1</td>
</tr>
<tr>
<td><em>Drd4</em></td>
<td>NP_031904.1</td>
<td>NP_037076.1</td>
</tr>
<tr>
<td><em>Gfap</em></td>
<td>NP_034407.2</td>
<td>NP_058705.2</td>
</tr>
<tr>
<td><em>Grm1</em></td>
<td>NP_058672.1</td>
<td>NP_058707.1</td>
</tr>
<tr>
<td><em>Grm2</em></td>
<td>NP_001153825.1</td>
<td>NP_001099181.1</td>
</tr>
<tr>
<td><em>Mmp2</em></td>
<td>NP_032636.1</td>
<td>NP_112316.2</td>
</tr>
<tr>
<td><em>Oprm1</em></td>
<td>NP_001034741.1</td>
<td>NP_037203.1</td>
</tr>
<tr>
<td><em>Penk</em></td>
<td>NP_001002927.1</td>
<td>NP_058835.1</td>
</tr>
<tr>
<td><em>Pmp22</em></td>
<td>NP_032911.1</td>
<td>NP_058733.1</td>
</tr>
<tr>
<td><em>Syn1</em></td>
<td>NP_038708.3</td>
<td>NP_062006.1</td>
</tr>
<tr>
<td><em>Timp-1</em></td>
<td>NP_001037849.1</td>
<td>NP_446271.1</td>
</tr>
</tbody>
</table>
Table 7: Mouse and rat orthologs of immune loci.

<table>
<thead>
<tr>
<th>Immune Locus</th>
<th>GenBank Accession No. (Mus)</th>
<th>GenBank Accession No. (Rattus)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apcs</td>
<td>NP_035448.2</td>
<td>NP_058866.2</td>
</tr>
<tr>
<td>CD2</td>
<td>NP_038514.1</td>
<td>NP_036962.1</td>
</tr>
<tr>
<td>CD8α</td>
<td>NP_113726.1</td>
<td>NP_001074579.1</td>
</tr>
<tr>
<td>CD8β</td>
<td>NP_033988.1</td>
<td>NP_113727.1</td>
</tr>
<tr>
<td>CD48</td>
<td>NP_031675.1</td>
<td>NP_620803.1</td>
</tr>
<tr>
<td>CD74</td>
<td>NP_001036070.1</td>
<td>NP_037201.1</td>
</tr>
<tr>
<td>CRP</td>
<td>NP_031794.3</td>
<td>NP_031794.3</td>
</tr>
<tr>
<td>IL2</td>
<td>NP_032392.1</td>
<td>NP_446288.1</td>
</tr>
<tr>
<td>IL3</td>
<td>NP_034686.2</td>
<td>NP_113701.1</td>
</tr>
<tr>
<td>IL4</td>
<td>NP_067258.1</td>
<td>NP_958427.1</td>
</tr>
<tr>
<td>IL9</td>
<td>NP_032399.1</td>
<td>XP_341488.2</td>
</tr>
<tr>
<td>Thy1</td>
<td>NP_033408.1</td>
<td>NP_036805.1</td>
</tr>
</tbody>
</table>

Table 8: Mouse and rat orthologs of essential loci.

<table>
<thead>
<tr>
<th>Essential Locus</th>
<th>GenBank Accession No. (Mus)</th>
<th>GenBank Accession No. (Rattus)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amh</td>
<td>NP_031471.2</td>
<td>NP_037034.1</td>
</tr>
<tr>
<td>Bax</td>
<td>NP_031553.1</td>
<td>NP_058755.1</td>
</tr>
<tr>
<td>Bdnf</td>
<td>NP_031566.4</td>
<td>NP_036645.1</td>
</tr>
<tr>
<td>Egr1</td>
<td>NP_031939.1</td>
<td>NP_036683.1</td>
</tr>
<tr>
<td>Esr1</td>
<td>NP_031982.1</td>
<td>NP_036821.1</td>
</tr>
<tr>
<td>Hoxa1</td>
<td>NP_034579.3</td>
<td>NP_037207.1</td>
</tr>
<tr>
<td>Lif</td>
<td>NP_032527.1</td>
<td>NP_071532.2</td>
</tr>
<tr>
<td>Mgat1</td>
<td>NP_034924.3</td>
<td>NP_110488.1</td>
</tr>
<tr>
<td>Pgr</td>
<td>NP_032855.2</td>
<td>NP_074038.1</td>
</tr>
<tr>
<td>Prlr</td>
<td>NP_035299.4</td>
<td>NP_001029283.1</td>
</tr>
<tr>
<td>Smn</td>
<td>NP_035550.1</td>
<td>NP_071954.1</td>
</tr>
<tr>
<td>Vegfa</td>
<td>NP_001020421.2</td>
<td>NP_114024.2</td>
</tr>
</tbody>
</table>
Table 9: Rates of nonsynonymous and synonymous substitution at imprinted loci. Coding sequences were obtained through the RefSeq database. Rates reflect the nonsynonymous and synonymous substitutions between mouse and rat orthologous sequences.

<table>
<thead>
<tr>
<th>Imprinted Locus</th>
<th>$d_N \pm$ s.e.</th>
<th>$d_S \pm$ s.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asb4</td>
<td>0.010 ± 0.003</td>
<td>0.263 ± 0.035</td>
</tr>
<tr>
<td>Cpa4</td>
<td>0.031 ± 0.006</td>
<td>0.233 ± 0.033</td>
</tr>
<tr>
<td>Dio3</td>
<td>0.003 ± 0.002</td>
<td>0.152 ± 0.028</td>
</tr>
<tr>
<td>Dlx5</td>
<td>0.005 ± 0.002</td>
<td>0.062 ± 0.017</td>
</tr>
<tr>
<td>Grb10</td>
<td>0.007 ± 0.002</td>
<td>0.235 ± 0.034</td>
</tr>
<tr>
<td>Igf2</td>
<td>0.015 ± 0.006</td>
<td>0.039 ± 0.017</td>
</tr>
<tr>
<td>Igf2r</td>
<td>0.032 ± 0.003</td>
<td>0.223 ± 0.012</td>
</tr>
<tr>
<td>Ins1</td>
<td>0.034 ± 0.014</td>
<td>0.262 ± 0.066</td>
</tr>
<tr>
<td>Ins2</td>
<td>0.025 ± 0.010</td>
<td>0.174 ± 0.052</td>
</tr>
<tr>
<td>Rasgrf1</td>
<td>0.039 ± 0.010</td>
<td>0.123 ± 0.035</td>
</tr>
<tr>
<td>Snrpn</td>
<td>0.002 ± 0.001</td>
<td>0.124 ± 0.026</td>
</tr>
<tr>
<td>Usp29</td>
<td>0.050 ± 0.005</td>
<td>0.274 ± 0.026</td>
</tr>
</tbody>
</table>

Table 10: Rates of nonsynonymous and synonymous substitution at non-essential loci. Coding sequences were obtained through the RefSeq database. Rates reflect the nonsynonymous and synonymous substitutions between mouse and rat orthologous sequences.

<table>
<thead>
<tr>
<th>Non-Essential Locus</th>
<th>$d_N \pm$ s.e.</th>
<th>$d_S \pm$ s.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ctnf</td>
<td>0.029 ± 0.009</td>
<td>0.094 ± 0.008</td>
</tr>
<tr>
<td>D2dr</td>
<td>0.001 ± 0.001</td>
<td>0.139 ± 0.022</td>
</tr>
<tr>
<td>Drd4</td>
<td>0.026 ± 0.005</td>
<td>0.178 ± 0.028</td>
</tr>
<tr>
<td>Gfap</td>
<td>0.014 ± 0.004</td>
<td>0.203 ± 0.030</td>
</tr>
<tr>
<td>Grm1</td>
<td>0.006 ± 0.002</td>
<td>0.213 ± 0.018</td>
</tr>
<tr>
<td>Grm2</td>
<td>0.007 ± 0.002</td>
<td>0.204 ± 0.019</td>
</tr>
<tr>
<td>Mmp2</td>
<td>0.007 ± 0.002</td>
<td>0.221 ± 0.025</td>
</tr>
<tr>
<td>Oprm1</td>
<td>0.023 ± 0.006</td>
<td>0.264 ± 0.035</td>
</tr>
<tr>
<td>Penk</td>
<td>0.014 ± 0.005</td>
<td>0.318 ± 0.052</td>
</tr>
<tr>
<td>Pmp22</td>
<td>0.014 ± 0.006</td>
<td>0.175 ± 0.043</td>
</tr>
<tr>
<td>Syn1</td>
<td>0.314 ± 0.021</td>
<td>0.534 ± 0.042</td>
</tr>
<tr>
<td>Timp-1</td>
<td>0.066 ± 0.013</td>
<td>0.168 ± 0.037</td>
</tr>
</tbody>
</table>
Table 11: Rates of nonsynonymous and synonymous substitution at essential loci. Coding sequences were obtained through the RefSeq database. Rates reflect the nonsynonymous and synonymous substitutions between mouse and rat orthologous sequences.

<table>
<thead>
<tr>
<th>Essential Locus</th>
<th>(d_N \pm s.e.)</th>
<th>(d_S \pm s.e.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amh</td>
<td>0.046 ± 0.006</td>
<td>0.240 ± 0.025</td>
</tr>
<tr>
<td>Bax</td>
<td>0.009 ± 0.005</td>
<td>0.037 ± 0.017</td>
</tr>
<tr>
<td>Bdnf</td>
<td>0.004 ± 0.002</td>
<td>0.076 ± 0.020</td>
</tr>
<tr>
<td>Egr1</td>
<td>0.008 ± 0.002</td>
<td>0.151 ± 0.020</td>
</tr>
<tr>
<td>Esr1</td>
<td>0.010 ± 0.003</td>
<td>0.232 ± 0.036</td>
</tr>
<tr>
<td>Hoxa1</td>
<td>0.018 ± 0.005</td>
<td>0.116 ± 0.023</td>
</tr>
<tr>
<td>Lif</td>
<td>0.044 ± 0.011</td>
<td>0.294 ± 0.055</td>
</tr>
<tr>
<td>Mga1</td>
<td>0.013 ± 0.004</td>
<td>0.177 ± 0.025</td>
</tr>
<tr>
<td>Pgr</td>
<td>0.037 ± 0.005</td>
<td>0.230 ± 0.019</td>
</tr>
<tr>
<td>Prlr</td>
<td>0.044 ± 0.006</td>
<td>0.205 ± 0.025</td>
</tr>
<tr>
<td>Smn</td>
<td>0.023 ± 0.006</td>
<td>0.296 ± 0.047</td>
</tr>
<tr>
<td>Vegfa</td>
<td>0.012 ± 0.004</td>
<td>0.084 ± 0.017</td>
</tr>
</tbody>
</table>

Table 12: Rates of nonsynonymous and synonymous substitution at immune loci. Coding sequences were obtained through the RefSeq database. Rates reflect the nonsynonymous and synonymous substitutions between mouse and rat orthologous sequences.

<table>
<thead>
<tr>
<th>Immune Locus</th>
<th>(d_N \pm s.e.)</th>
<th>(d_S \pm s.e.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apcs</td>
<td>0.111 ± 0.017</td>
<td>0.187 ± 0.039</td>
</tr>
<tr>
<td>CD2</td>
<td>0.109 ± 0.013</td>
<td>0.251 ± 0.035</td>
</tr>
<tr>
<td>CD8a</td>
<td>0.175 ± 0.020</td>
<td>0.266 ± 0.046</td>
</tr>
<tr>
<td>CD8β</td>
<td>0.128 ± 0.018</td>
<td>0.245 ± 0.047</td>
</tr>
<tr>
<td>CD48</td>
<td>0.194 ± 0.022</td>
<td>0.201 ± 0.038</td>
</tr>
<tr>
<td>CD74</td>
<td>0.123 ± 0.020</td>
<td>0.282 ± 0.054</td>
</tr>
<tr>
<td>CRP</td>
<td>0.137 ± 0.018</td>
<td>0.187 ± 0.039</td>
</tr>
<tr>
<td>IL2</td>
<td>0.085 ± 0.016</td>
<td>0.162 ± 0.043</td>
</tr>
<tr>
<td>IL3</td>
<td>0.278 ± 0.035</td>
<td>0.207 ± 0.045</td>
</tr>
<tr>
<td>IL4</td>
<td>0.253 ± 0.033</td>
<td>0.339 ± 0.072</td>
</tr>
<tr>
<td>IL9</td>
<td>0.146 ± 0.023</td>
<td>0.258 ± 0.061</td>
</tr>
<tr>
<td>Thy1</td>
<td>0.100 ± 0.020</td>
<td>0.170 ± 0.039</td>
</tr>
</tbody>
</table>